

# THIS WEEK



## EDITORIALS

**MARS** The continuing search for the secrets of the red planet **p.446**

**WORLD VIEW** Science must come to terms with life after the crash **p.447**

**LATTICE** Lighter than air material could deaden future sounds **p.448**

## Breeding contempt

*By failing to explain why a moratorium on breeding chimpanzees seems not to have been enforced, the US National Institutes of Health risks a further loss of public support for chimp research.*

Research using chimpanzees is under closer scrutiny than usual in the United States, with the Institute of Medicine scheduled to report next month on whether government funding for the work continues to be scientifically justified. So it is, to put it mildly, an unfortunate time for questions to be raised about a key research centre's ability and willingness to follow a 16-year-old moratorium on chimpanzee breeding. It is also a sensitive time for the National Institutes of Health (NIH) in Bethesda, Maryland — the agency that funds the research on behalf of the US taxpayer — to be reluctant to answer key questions about its enforcement of the moratorium at the New Iberia Research Center (NIRC), near Lafayette in Louisiana.

The centre houses 348 chimps, of which 117 are NIH-owned, and receives a steady stream of public funds for their upkeep — around US\$1 million a year. The NIRC has also received more than \$6 million since 2002 to supply very young chimps to the NIH's National Institute of Allergy and Infectious Diseases for hepatitis C studies — a project that the institute calls “essential” for progress towards a vaccine.

The moratorium is designed to cap upkeep costs and prevent the taxpayer from picking up the bill for a lifetime of care for chimps born to publicly owned animals. But as *Nature* reports this week (see page 453), the NIRC has been forced to admit to the births of 137 infant chimps to NIH-owned animals between 2000 and 2009 — revealed by the Humane Society of the United States in Washington DC using state and federal laws on freedom of information.

In defence of the NIRC, its director Thomas Rowell points to a clause in the written agreement between the NIH and the NIRC, which stipulates that infant chimps born during the funding period become the property of the centre, not of the NIH. The taxpayer is thus not responsible, he argues, and a valuable biomedical resource is maintained.

Yet surely the provision was written to protect the NIH from financial responsibility for infant chimps that are born occasionally and accidentally — not as the go-ahead for a breeding programme conducted with a nod and a wink? After all, the agreement also stipulates that “as a condition of this award, a moratorium on breeding activities... will remain in effect”.

The moratorium was introduced by the NIH in 1995 for fiscal reasons; it costs at least \$300,000 to support a single chimp for its lifetime. But today the moratorium carries increasing moral weight, as public opposition to chimp research has grown and the United States has become one of only two nations that supports such work.

Almost as disturbing as the NIRC's stance — which seems cavalier at best, and openly defiant of the moratorium at worst — is the silence with which the NIH has greeted repeated enquiries from *Nature* and others. It is not as if the agency has been caught off-guard; it has known that the issue was simmering since the summer of 2010, when the Humane Society took evidence of the 137 chimp births to the US Senate committee that funds the NIH. In a written report, the

committee asked the NIH to look into the allegation that the NIRC was openly breaching the moratorium. The NIH, responding in writing, minimized the findings, saying that only 28 births to NIH-owned animals had been recorded at the centre since 1995, and arguing that the NIRC “is in compliance with the conditions of the moratorium” because it assumed ownership of the animals.

The agency has since refused to discuss its responses and has not answered a petition from the Humane Society that was filed eight months ago with the Department of Health and Human Services, the NIH's parent agency, asking that the centre's NIH funding be revoked.

The US public, which supports biomedical research in good faith and is increasingly concerned about the welfare of research animals in general, and chimps in particular, deserves better. The NIH should fully explain how it enforces the breeding moratorium at the NIRC. It should say how it is possible that 137 births in a decade were somehow permitted. And it should address publicly how one of its institutes can in good conscience pay for and use a steady supply of infant chimps born in apparent breach of the moratorium.

Until officials are fully open with the public and demonstrate convincingly that the research centres they help to fund are both competent and compliant with the rules — and that, when they are not, the proper sanctions will ensue — US chimpanzee research risks losing public support entirely, and with good reason. ■

**“The US public supports biomedical research in good faith and deserves better.”**

## The heritage trail

*Egypt and Libya can look to the past to help build a more stable future.*

The streets of Cairo burned again this week and Tahrir Square swelled once more with protestors — proof, if any were needed, that the Arab Spring will take time to reach full bloom. Uncertainty is a feature of revolutions, as those who have just sealed their own victory in Libya will be well aware. And against such a background, the interests of archaeologists in both countries are, naturally, low on the list of priorities.

Yet the rich archaeological heritage of Egypt and Libya could contribute to their rebuilding. Cultural remains there are so extensive that both nations are effectively open-air museums. Egypt's sites offer a window onto an ancient civilization that was stable for some 3,000 years. Libya's multicultural archaeology is unique, with Punic and

Roman sites to the west, Greek and Egyptian to the east, and Berber to the south. And its boundless desert is home to one of the world's largest and earliest collections of prehistoric rock art, as well as to clues to early Saharan civilizations and human movement between Africa and Europe (see *Nature* <http://dx.doi.org/10.1038/news.2011.132>; 2011).

For foreign and local archaeologists unable to return safely to work on such sites, the present instability is a source of extreme frustration. Yet they seem reticent to start pressing for archaeology and cultural heritage to be given attention. The revolutions in both countries mark a chance for a radical change of past structures and practices. Revitalizing archaeology in Egypt and Libya will not happen immediately, but it is not too soon to start planning. When the security situation allows foreign researchers to return to Egypt and Libya, funding agencies must be ready to boost collaborations between them and their colleagues in the host nations.

In Egypt, as the News Feature on page 464 shows, the picture for researchers was volatile even before the recent upsurge in violence. The challenge now is to get moving again what was a reasonably well functioning national archaeological service. The tourists will not be back in numbers for years to come, depriving Egypt's archaeology of precious funds. But given the nation's solid archaeology base, there are few reasons to be pessimistic about its long-term prospects.

By contrast, Libya's infrastructure is nowhere near commensurate with the size and diversity of its cultural heritage (see *Nature* <http://dx.doi.org/10.1038/nature.2011.9396>; 2011). Libyan researchers hope that the revolution will bring about a sea change after Gaddafi's 42-year regime, which viewed archaeology as a vestige of colonialism.

To garner public and political support, education on the significance of Libya's heritage is needed at all levels, from schools, to stakeholders living and working near sites, to politicians and business people. Plans must be drafted to reform and develop the country's archaeology, and to train abroad a corps of young archaeologists in modern techniques.

Until then, foreign researchers' input will be essential. They can also play a key part in helping to counter the immediate threat to Libya's heritage posed by the likely reconstruction boom in housing,

***“Revitalizing archaeology will take time, but it is not too soon to start planning.”***

infrastructure and the oil industry. Libya urgently needs to be surveyed to map all of its heritage sites accurately, catalogue their contents in databases, and then protect the densest concentrations with national parks. Its archaeological services have none of this

— when NATO asked for GPS coordinates for the most important sites, to avoid them in airstrikes, no list existed and researchers worldwide scrambled to assemble a rough guide.

Libya's revolution marks a diplomatic success for NATO, and for France, the United Kingdom and the United States in particular. If researchers put forward sensible proposals for cultural management, they may get more political support from those international politicians who helped topple Gaddafi than they think. Oil firms operating in Libya should also fund projects and help to preserve sites, as some already are by sharing their maps.

The new leaders in Libya and Egypt inherit nations in flux and need to establish shared goals with their people. A sense of pride in developing and protecting cultural heritage may be one small step towards that. ■

## Seeing red

*Latest mission to Mars promises close-up view of planet's surface.*

For generations, people gazing up at Mars could only mythologize the pink sparkle in the sky. Only in the seventeenth century did telescopes pointed at the planet reveal an orb, capped in white at the poles, with a mysterious dark splotch around its middle. Centuries on still, it took a barrage of spacecraft at the dawn of the space age to prise its secrets open: ice caps that contained not just water but frozen carbon dioxide, and Syrtis Major, a dark field of lava squeezed out by long-dead volcanoes, which seemed to be regularly scoured by wind. More recent orbiters, reaching their peak with the spy-camera resolution of the Mars Reconnaissance Orbiter, have caught Mars in action today — red-handed, if you like. Streaks on crater walls that come and go with the seasons are probably the signature of near-surface water. And little flashes of colour appear as if from nowhere as tiny asteroids penetrate and expose the blue ice below.

Mike Malin, a scientist and businessman in San Diego, California, has been a constant witness to this steady march of knowledge. Indeed, as the News Feature on page 460 makes clear, he made much of it happen, having had a hand in building the cameras on every NASA Mars orbiter that followed the 1975 Viking missions. But his cameras have not seen the surface before. Now finally, with the launch of the Mars Science Laboratory, or Curiosity, scheduled for as early as 25 November, Malin's cameras will, he hopes, at last view Mars in close-up. One of his camera systems sits at the end of the new rover's robotic arm, a microscopic imager with the power to see, for the first time, tiny grains of silt thought to exist on Mars — particles that would bear the history of billions of years of erosion from wind and water.

This is one benefit of a coordinated and sequenced NASA Mars programme. But at what cost? Getting to Mars has never been cheap. Just ask Russia, which spent US\$163 million to launch on 8 November

an attempt to retrieve a few cups of soil from one of Mars's moons. Now the stricken spacecraft is stuck in Earth orbit, laden with fuel: cheap as failed Mars missions go, but dearly expensive for a firework. NASA's Curiosity, at \$2.5 billion, is at the luxury end of the spectrum. Part of this staggering cost is because of the checks and further checks that NASA hopes will guarantee the mission's success. But it also reflects the sophistication of the science payload, which will investigate whether ancient watery environments may have been suitable for life.

Still, NASA has largely danced around the question of whether it will pursue Martian life itself. Last week, before the US Congress, Mars exploration advocates were explaining their next steps — a joint plan with the European Space Agency — and how this question would finally be addressed. A planned orbiter in 2016 would sniff for trace amounts of methane, a potential biosignature. And then in 2018 a rover would land and drill to see if microorganisms have been living beneath the surface. This rover would also grab bits of rock and soil as the first stage in a three-pronged plan to bring samples back to Earth, a long-term priority for Mars scientists.

But sample return of any sort is expensive. One credible assessment from the National Academies' decadal survey review puts the costs of a three-stage Mars sample return at as much as \$8.5 billion. That is why budget minders for the administration of President Barack Obama are rightly concerned about embarking down this road — \$8.5 billion happens to be roughly the price tag for a similarly ambitious mission, the James Webb Space Telescope, a successor to the Hubble Space Telescope. But the Webb telescope at least buys diversity: the ability to investigate millions of galaxies and stars, and even to look for the atmospheric biosignatures of planets not too much larger than Earth. A Mars sample return buys you a few rocks from one specific place. After Viking, in the late 1970s and early 1980s, NASA — exploring space during a terrestrial economic slump and a time of post-

Apollo drift — did not launch a Mars mission for almost a generation. And that is why Curiosity, for all its prowess, may also be the last in the current line of landed robotic Mars missions. Godspeed, Curiosity. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunq](http://go.nature.com/xhunq)



## Science's attitudes must reflect a world in crisis

Colin Macilwain welcomes a wake-up call to science policy-makers to address the consequences of current political and economic upheavals.

Those involved in science policy sometimes seem to me to be sleep-walking through the greatest crisis to afflict the West since the Second World War. True, from the point of view of the scientist at the bench, grants continue to flow and results continue to be published. Perhaps this is why wider discourse about science's role in society has hardly budged an inch.

For the past three years, I have grown steadily more impatient with this 'business as usual' approach. Whenever an academy president or research chief stands up to speak in public, I have been waiting for them to explain how they will do things differently. They never do.

At the World Science Forum in Budapest last week, some scientific leaders finally acknowledged the new reality. In particular, representatives of developing countries — which account for a fast-growing share of global science — talked of radical reorientation of research priorities to better match the pressing needs of their populations. And behind the scenes, analysts are mapping out fascinating, and sometimes alarming, possible scenarios for global science after the crash.

Challenged by Hungarian Prime Minister Viktor Orbán to face up to the implications of the economic crisis, prominent Western representatives at the forum, such as William Colglazier, science adviser to US secretary of state Hillary Clinton, and Chris Llewellyn Smith, former chief of Europe's particle-physics laboratory CERN, failed to do so in their plenary talks. And some speakers were clearly more comfortable discussing the planet's ecological crises than the economic ones currently alarming the general population.

Questions were soon raised, however, when Princess Sumaya bint el Hassan of Jordan's Royal Scientific Society captured the mood of the developing world. "We must ask ourselves why so much scientific research is driven by the consumer needs of a tiny elite," she said. "We're being naive if we envisage business-as-usual for science in the new century."

Some things remained unsaid at Budapest: no one criticized science's failure to join engineering groups in highlighting the lack of solid, productive foundations for the two-decades-long boom that ended abruptly in 2008. Nor was there criticism of failure to expose the pseudoscience that underpinned the exotic financial instruments that played such a central part in both the boom and the bust.

But the forum did reveal the beginnings of a serious response by scientific leaders to the tumult ahead. Despite much cosy rhetoric about defending research funding, one uncomfortable but realistic scenario is for it to nose-dive, perhaps by one-third in the United States and the United Kingdom in real terms, over the next five years. Even though other nations

will spend more money, that sort of change will wreak havoc.

One great danger is that scarce funding will consolidate around single-discipline research — even though everyone knows that the most valuable work is now multidisciplinary. An associated danger, already revealed in the US Congress, is that the social sciences will be expelled from the temple — just when, as Llewellyn Smith pointed out in Budapest, the hard sciences need to invite them in to help public engagement.

But the political outcomes of the crisis aren't yet clear enough to enable scientists to plan around them. That has led the International Council for Science (ICSU) in Paris, the global association of academies and scientific societies, to conduct a foresight exercise that explores how science as a whole might change shape over the next 20 years.

An ICSU task force led by physicist John Marks has been looking at all the drivers of global science and has consolidated them into two overriding forces: engagement with society and globalization (as opposed to nationalism). Plotting these two against each other, Marks told the forum, produces four distinct scenarios for the future — whatever the level of funding.

The first and most sunny, with more globalization and high engagement, would see a series of positive outcomes, including much more interdisciplinary research. The second — more globalization but low engagement — is rather like what we had before the crash, only worse. The ICSU PowerPoint slide for this showed bunches of vainglorious yuppies with mobile phones and portable computers, doubtless creating more gizmos and expensive drugs that most people in the world can't afford. The third scenario would have

more nationalism, with high engagement. That might create a series of little Denmarks pulling away from each other to deal with their own problems, with their own research strategies and regulatory regimes.

Finally, and most ominously, there's more nationalism, with less engagement. This predicts old-fashioned, stick-to-your-knitting, single-discipline science, aligned with resurgent nationalism. The slide for this one had a mushroom cloud at one stage, but Marks settled for a barely more reassuring image of some darkly lurking battleships.

The ICSU exercise isn't complete yet and not everyone sees its value. But it outlines the choices that science faces. Scientists have always cultivated globalization, and can keep pushing for it. Engagement is different; a tribal disdain for the social sciences still holds sway in the laboratory, as does a haughty disregard for the views and demands of the general public. Both outlooks need to be jettisoned if science is to contribute and thrive in this new world. ■

Colin Macilwain is a contributing correspondent with Nature.  
e-mail: cfmworldview@gmail.com

ANALYSTS ARE  
MAPPING OUT  
**FASCINATING**,  
AND SOMETIMES  
**ALARMING**,  
SCENARIOS FOR  
GLOBAL SCIENCE  
AFTER THE CRASH.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/hlaaos](http://go.nature.com/hlaaos)



# SEVEN DAYS

The news in brief

## POLICY

### US budget cuts

Congressional legislators have managed to pass a portion of the US government 2012 budget, including allocations for NASA, the National Science Foundation and the Food and Drug Administration, among other science-related agencies. The 17 November settlement was not entirely bad for those bracing for deep cuts — with the exception of the White House Office of Science and Technology Policy, whose budget was slashed by one-third. But all science agencies face greater pain from January 2013, after legislators couldn't agree by 23 November on how to trim the federal deficit by \$1.2 trillion, a failure that triggers across-the-board budget cuts. See page 455 for details.

### Avastin revoked

The US Food and Drug Administration has withdrawn its approval of the drug Avastin (bevacizumab) to treat advanced breast cancer. The agency announced almost a year ago that it would revoke approval because the drug's risks outweigh its benefits. Avastin's maker, Genentech in San Francisco, California, challenged that verdict, but the agency finalized its decision on 18 November. See [go.nature.com/46m7w5](http://go.nature.com/46m7w5) for more.

### Italian shake-up

Electrical engineer Francesco Profumo will be minister for education, universities and research in Italy's emergency administration, which was announced on 16 November by Prime Minister Mario Monti. Profumo was rector of the Polytechnic of Turin until August, when he was made president of Italy's National Research Council funding agency. That appointment

was part of a broader initiative to make the Italian research system more transparent and meritocratic (see *Nature* 476, 386; 2011); Profumo will now oversee these reforms. There are no politicians in Monti's cabinet, which is intended to govern until elections in 2013.

### Antibiotic alert

The European Commission has unveiled a five-year strategy to tackle the growing problem of antibiotic resistance, urging greater prudence in the use of antimicrobials in human and veterinary medicine. The action plan, which recommends strengthening surveillance systems to better track and report cases of

resistance, was published on 17 November. On the same day, the European Centre for Disease Prevention and Control in Solna, Sweden, published statistics showing how bacteria resistant to the carbapenems, a key last-line-of-defence class of antibiotic, are spreading through Europe. See [go.nature.com/zn3csi](http://go.nature.com/zn3csi) for more.

## BUSINESS

### Gilead spends big

Gilead Sciences in Foster City, California, is spending US\$11 billion to acquire small biotechnology firm Pharmasset, it announced on 21 November. Pharmasset,

three-dimensional replica of the site. (Vincent Rossi from the Smithsonian's 3D digitization programme is pictured.) The team had to finish by December because the area will soon become a highway; the fossils were first spotted by a road-building firm working on the site in 2010. See [go.nature.com/kd24pr](http://go.nature.com/kd24pr) for more.



N. D. PYENSON

## Scanning whales in the desert

of Princeton, New Jersey, has three treatments for hepatitis C in clinical trials; Gilead also has interests in this area. At \$137 per share, the agreed price was an 89% premium on Pharmasset's closing price at the end of the previous week.

## RESEARCH

### Faster than light 2

The OPERA experiment, which made headlines in September for saying that neutrinos could travel faster than light, released further data supporting its claim on 17 November. The experiment measures the speed of a pulsed beam of neutrinos sent about



R. TABASSUM/AFP/GETTY IMAGES

730 kilometres from CERN, Europe's particle-physics lab near Geneva, Switzerland, to Gran Sasso National Laboratory near L'Aquila, Italy. In the latest run, the pulses were shorter and more precisely timed, ruling out some sources of systematic error. But physicists are likely to remain doubtful until an independent experiment replicates the finding. MINOS at Fermilab in Batavia, Illinois, hopes to weigh in early in 2012. See [go.nature.com/3f8fpr](http://go.nature.com/3f8fpr) for more.

## Higgs hunting

Data from experiments at the Large Hadron Collider have further restricted the range of possible masses for the elusive Higgs boson — the particle thought to give other fundamental particles their mass. On 18 November, research teams from the particle accelerator at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, revealed the latest results, which restrict the Higgs to a mass of 114–141 gigaelectronvolts. Data analysis scouring this final hiding place should be complete by the end of this year. See page 456 for more.

## Extreme weather

Climate change will make extremes of hot weather more frequent and severe



in coming decades, the Intergovernmental Panel on Climate Change warned in a report released on 18 November. But the effect on patterns of rainfall, flood risk and tropical storms is uncertain, said the report. It is not clear, for example, what will happen to the frequency of localized weather events such as last year's catastrophic floods in Pakistan (**pictured**). Last week's report was a summary for policy-makers; the full version is scheduled for release early next year. See [go.nature.com/nxyavf](http://go.nature.com/nxyavf) for more.

## PEOPLE

### Europe's adviser

Europe's first chief scientific adviser (CSA) will be Anne Glover, a molecular and cell biologist who is currently CSA to the Scottish government, sources told *Nature* on 21 November. Glover's

appointment comes more than two years after José Manuel Barroso, the president of the European Commission, pledged on 15 September 2009 to create the post. But the details of her role are still unclear, including how much power and freedom she will have in providing advice and influencing policy-making. See [go.nature.com/kvmwya](http://go.nature.com/kvmwya) for more.

## NASA science head

John Grunsfeld, an astrophysicist and astronaut who fixed the Hubble Space Telescope, has been chosen as associate administrator of NASA's science mission directorate, according to sources with knowledge of the selection. Grunsfeld is currently deputy director of the Space Telescope Science Institute in Baltimore, Maryland, which operates Hubble. He would replace Ed Weiler, who resigned in September. See [go.nature.com/u929zt](http://go.nature.com/u929zt) for more.

## Stolen data?

The scientist behind a partially retracted paper that linked chronic fatigue syndrome to a virus, XMRV, has been arrested and jailed in relation to a lawsuit brought by her former employer. Judy Mikovits had been research director at the Whittemore Peterson Institute for

## COMING UP

**26 NOVEMBER**  
NASA's Mars Science Laboratory is scheduled to launch.

[go.nature.com/gmbdai](http://go.nature.com/gmbdai)

**28 NOV–9 DEC**  
International climate negotiations continue in Durban, South Africa, with the 17th meeting of parties to the United Nations Framework Convention on Climate Change. A key debate will be the future of the Kyoto Protocol, which expires next year (see page 454).  
[www.cop17-cmp7durban.com](http://www.cop17-cmp7durban.com)

**30 NOVEMBER**  
Patent protection runs out on Pfizer's cholesterol-lowering drug Lipitor (atorvastatin), the world's top-selling prescription medicine.

Neuro-Immune Disease in Reno, Nevada, but was fired barely two months ago. On 18 November she was arrested by police in Ventura County, California; the lawsuit claims she absconded with laboratory notebooks and proprietary information. See [go.nature.com/uwlzph](http://go.nature.com/uwlzph) for more.

## WHO retains chief

Margaret Chan, who has been director-general of the World Health Organization for the past five years, looks set to retain her post for another five-year term. The agency said on 21 November that Chan was the only candidate nominated for the position by member states. Its executive board meets in January 2012 to vote on proposing Chan for re-election: an appointment will be formally decided in May by the World Health Assembly.

**NATURE.COM**  
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

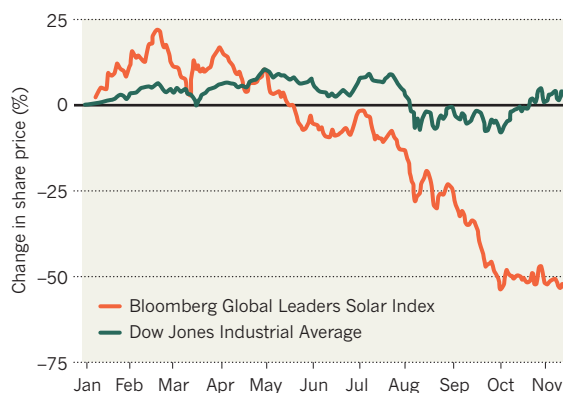
SOURCE: BLOOMBERG

## TREND WATCH

Slow demand for solar panels, oversupply of products and materials, and subsidy cuts are biting into the industry's profits. Last week, Germany's biggest solar-panel maker, SolarWorld in Bonn, reported its first quarterly net loss in two years, and the chief financial officer of panel maker Q-Cells in Bitterfeld-Wolfen resigned after a second quarterly loss. Even Chinese panel maker JinkoSolar in Shanghai and polysilicon manufacturer Daqo in Chongqing City said they had missed sales targets this quarter.

### SOLAR STOCKS KEEP FALLING

Solar-panel makers and materials suppliers continue to struggle in 2011 as cheap photovoltaic modules oversupply the market.

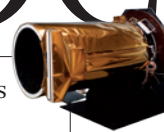


# NEWS IN FOCUS

**FUNDING** A budget reprieve for some of the US science agencies **p.455**

**PHYSICS** Few hiding places now remain for the Higgs particle **p.456**

**ETHICS** Synthetic biologists debate production of 'green explosives' **p.458**



**SPACE** The vision that guides the new Mars rover **p.460**

T. MUELLER/THE NEW YORK TIMES/REDUX/EYEVINE



Infant chimpanzees born recently at the New Iberia Research Center in Louisiana.

POLICY

## Lab bred chimps despite ban

*US facility says it broke no rules — but opponents cry foul.*

BY MEREDITH WADMAN

**T**he largest and most high-profile chimpanzee research centre in the United States has acknowledged to *Nature* that 137 infant chimpanzees have been born to federally owned animals under its care since 2000, despite a government moratorium on such births. The centre says that it abided by the policy because the infant chimps are not

supported by federal funds. But critics claim that any breeding of government chimps violates the spirit of the ban, which they consider was partly a response to ethical concerns about research on chimps. The United States and Gabon are the only countries whose governments are known to allow invasive chimpanzee research.

Thomas Rowell, director of the New Iberia Research Center (NIRC) in Louisiana, has told

*Nature* that he does not dispute allegations that 123 chimps born at the centre between 2000 and 2009 — more than one-third of the total number currently housed there — have at least one parent owned by the National Institutes of Health (NIH) in Bethesda, Maryland. The NIH brought in a breeding moratorium in 1995 that has been included in agency contracts made with chimp research facilities ever since. But Rowell says that the NIRC, which is part of the University of Louisiana at Lafayette, is not in breach of the policy because, as specified in its written agreement with the NIH, the centre assumes ownership of the infants and pays for their care.

Details of the NIRC's breeding activities were first made known to the US Senate in 2010 when the Humane Society of the United States used Louisiana and US freedom of information laws to compile birth records of chimps born at the centre during the previous decade. The investigation showed that 137 of the animals had at least one federally owned parent. Attacks by adults killed 14 of the infant chimps; of the 123 that survived, the Humane Society's data indicate, 7 are owned by the NIH and not by the chimp centre, as the agreement requires. Rowell will not discuss the seven because he says the NIH is investigating the matter.

"They are breeding in what seems to be violation of government contracts," says Kathleen Conlee, the society's senior director of programme management, who notes that the centre receives about US\$1 million per year for the upkeep of federally owned chimps. "They shouldn't be receiving funds if they are not following the grant agreement." In March, the Humane Society petitioned the US Department of Health and Human Services — the NIH's parent agency — to revoke funding to the NIRC, and asked the Department of Justice to fine the centre as much as \$30 million.

After the society's findings were reported online last week by *Wired*, Rowell told *Nature* that the centre's actions are "consistent" with the terms of the award. The NIH will neither confirm nor deny that the NIRC is in violation of its agreement.

Gregory Kaebnick, an ethicist at the Hastings Center in Garrison, New York, says that even if the NIRC did take ownership of all 123 infant chimps, its reading of the ban "may be within ►

► **NATURE.COM**  
For more on chimp research in the United States see:  
[go.nature.com/ugfawq](http://go.nature.com/ugfawq)



► the letter of the law but well outside the spirit of the law". The moratorium inevitably reflects ethical objections to breeding, he argues, not just its costs. Requiring the NIRC to take responsibility for offspring of federally owned chimps is a reasonable way to cover accidental births due to the failure of contraceptive methods, but it should not be taken as a licence to breed, he adds.

Rowell says that the NIRC has taken several measures to reduce the number of animals being born compared with when it was "in a breeding mode" before 1995. Then, 20–40 births per year were typical. The measures include the use of intra-uterine devices in females; discontinuing immediate weaning of offspring, so that females remain infertile for up to four years after giving birth; vasectomizing selected males; and housing males and females separately where possible.

But because of the importance of maintaining breeding capacity in case the moratorium is lifted, Rowell says, the centre chose not to sterilize all animals. "The NIRC has made a good-faith attempt to maintain our colony as a high-value resource for both immediate and future need to the biomedical research community," he says. "The additional animals play a pivotal role in addressing global human health problems."

Although the NIH states that the moratorium was instituted for economic reasons, "more and more people are becoming conscious of the fact that chimps in research suffer in horrible ways that we don't want them to be suffering," says Bruce Wagman, a partner and animal-law expert in the San Francisco office of the law firm Schiff Hardin.

In mid-December, a committee of the US Institute of Medicine is due to opine on whether the NIH should continue to fund chimp research at all. The NIH called for the study after strong protests against its efforts to move 186 semi-retired chimpanzees back into active research (see *Nature* **467**, 507–508; 2010).

Paradoxically, the NIH has benefited from the continued breeding of federal animals. The agency's National Institute of Allergy and Infectious Diseases has paid the Louisiana centre more than \$6 million since 2002 to provide it with 4 to 12 infant chimpanzees annually for studies of hepatitis C and other viral diseases. According to the data gathered by the Humane Society, 50 of these chimps were born to one or more NIH-owned animals.

"Rather than passing off the costs and therefore saying 'it's no problem', the NIH should be doing everything it can to enforce the breeding moratorium," says Wagman. "The moratorium is absolute." ■

SEE EDITORIAL P.445



P. KITTIVONGSAKUL/AFP/GETTY

Nations that are vulnerable to flooding and drought stand to lose out if climate talks fail.

#### CLIMATE POLICY

# Showdown nears for climate deal

*Expected debate over Kyoto Protocol threatens to stall progress on other fronts.*

BY JEFF TOLLEFSON

There's a storm brewing in South Africa, although not the kind that climate scientists warned of last week in a report linking extreme weather to global warming (see <http://dx.doi.org/10.1038/nature.2011.9397>; 2011). This one is the political tempest that is swirling around the Kyoto Protocol. Forecasts suggest that it will hit in full force when delegates to the seventeenth meeting of the United Nations Framework Convention on Climate Change gather in Durban on 28 November — and it could threaten progress on an alternative agreement.

Negotiated in 1997, Kyoto combines commitments by wealthy nations (excluding the United States, which has never ratified the agreement) to limit greenhouse-gas emissions with monetary and technical support for poor nations. The commitments are set to expire next year and nothing comparable has materialized to take their place.

The world has changed since Kyoto was negotiated. Powerhouse developing economies are now among the world's leading

carbon dioxide emitters. Today, China holds the number one spot and India is fourth behind the United States and the European Union (EU). With Western nations in the midst of a financial crisis, there is little enthusiasm for renewing Kyoto, or for the underlying assumption that the developed world should shoulder the full burden of controlling greenhouse-gas emissions. In Durban, supporters of Kyoto — mainly developing nations — will make their last stand.

"The world can't punt this issue any longer," says Robert Stavins, an environmental economist at Harvard University's John F. Kennedy School of Government in Cambridge, Massachusetts. Stavins believes that negotiators will find some way to keep the Kyoto Protocol alive, if only in spirit, but he acknowledges that the meeting could also fall apart. "Some people would say that this is the death that is required

in order to have a rebirth" in the negotiating process, he says.

Cracks began to show in the framework's political underpinnings two

➔ **NATURE.COM**  
Read a Comment  
on the future of the  
Kyoto Protocol:  
[go.nature.com/c3zfk2](http://go.nature.com/c3zfk2)



years ago in Copenhagen (see *Nature* **462**, 966–967; 2009), when world leaders first broke with the Kyoto formula to offer the developing world rewards without commitments. Japan, Canada and Russia, which all agreed to the original protocol, said that they opposed a second commitment period. Even the EU may not support an extension of Kyoto, Stavins says, although its trail-blazing carbon trading system could continue even if the treaty becomes dormant.

In Copenhagen, and again last year in Cancún, Mexico, countries such as China, Brazil and South Africa joined wealthy nations in making voluntary commitments to limit emissions of greenhouse gases. In exchange, wealthy countries agreed to ramp up financing to help the poorest countries prepare for global warming. The framework that emerged is more inclusive than Kyoto, but lacks the formal stature of a fully binding international treaty. Developing countries, as a bloc, are reluctant give up on the old protocol until something more concrete takes its place.

“The call to mandate a new treaty in place of the Kyoto Protocol should be understood for what it is — rich countries backtracking on inconvenient commitments, at the expense of the poor and the planet,” says Lim Li Lin, who works with developing countries at the Third World Network, a non-profit advocacy group based in Penang, Malaysia.

Under the Copenhagen Accord, countries agreed to try to limit global warming to 2°C and to consider lowering that target to 1.5°C. Actual commitments to greenhouse-gas cuts fell well short of what studies suggest are needed to meet the 2°C goal. The accord also includes provisions to slow deforestation, promote adaptation and hasten the spread of green technologies, with wealthy countries promising US\$30 billion in financing by 2012, which will increase to \$100 billion annually by 2020.

Negotiators were able to advance many of these initiatives in Cancún last year, but crucial details regarding which institutions will manage this money, who will decide how it is spent and how emissions commitments will be registered and tracked remain unresolved. “Durban needs to put the entire architecture of climate-resilient development on a more solid basis,” says Koko Warner, a climate policy expert with the United Nations University Institute for Environment and Human Security in Bonn, Germany.

But many fear that frustration over Kyoto could stall progress in other areas of the talks and threaten the viability of the Copenhagen Accord.

“I think what really kept developing countries at the negotiating table in Copenhagen was the \$100 billion per year,” says Timmons Roberts, who heads the Center for Environmental Studies at Brown University in Providence, Rhode Island. “If it’s just another empty aid promise, the Copenhagen Accord is going to fall apart.” ■

## FUNDING

# US science agencies dodge deep cuts

*Partial 2012 budget mostly spares research, but deficit-reduction measures still pose a looming threat.*

BY IVAN SEMENIUK

**T**he most fractious and combative US Congress in recent memory is getting on with approving a 2012 budget — although perhaps only so that it can move more swiftly to the next battlefield. On 17 November, legislators passed a spending bill that includes allocations for several key science-related agencies. The bill has since been signed off by President Barack Obama.

The budget was a relief for researchers and their advocates, who had feared deeper cuts to science. Under the provisions of the bill, the National Science Foundation (NSF), the National Oceanic and Atmospheric Administration (NOAA), the National Institute of Standards and Technology and the Food and Drug Administration (FDA) all received modest increases (see table).

Even NASA, which has seen its budget shrink by nearly US\$1 billion over the past two years — essentially because it has ended the space-shuttle programme — kept all its major science initiatives. It even got an increase for the 6.5-metre James Webb Space Telescope (JWST), which has been subject to chronic cost overruns that at one point prompted deficit-focused Republicans to propose dropping it from the space agency’s budget altogether.

“That wakes you up,” says Kevin Marvel, executive officer of the American Astronomical Society in Washington DC. The JWST has now been restored, with continued support endorsed by both the House of Representatives and the Senate, and NASA’s science budget has increased by 3.1%. Although this by no means solves all of NASA’s budget woes — the agency is also facing questions about a series of missions to Mars (see Editorial, page 446) — it is a reassuring outcome in financially uncertain times. “Compared to where we could have been, we are in a good place for 2012,” says Marvel.

The NSF is also cautiously optimistic: its budget grew unexpectedly by almost US\$200 million dollars to just over \$7 billion. However, this falls short of the nearly \$7.8 billion that Obama requested for the NSF in February. That amount would have helped to maintain a ten-year

doubling of funding for the agency, as set out by the America COMPETES Act that was reauthorized last year.

The budget “lifts a huge cloud that was hanging over the agency and makes it into a much smaller one,” says Michael Lubell, director of public affairs for the American Physical Society in Washington DC. Having used economic-stimulus money to increase funding to researchers in 2009–10, the NSF was already expecting to scale back its grants next year. The 2012 budget is not large enough to avoid that altogether, but the agency should be able to maintain its commitments to major research facilities while seeing a smaller reduction in grant-acceptance rates.

The FDA received a \$50-million increase in its congressional appropriation, \$39 million of which will go towards implementing a food-safety law passed in January. “It’s

good that they saw the need to pass that funding, but we would like to have seen more,” says David Plunkett, a senior staff attorney at the

Center for Science in the Public Interest, an advocacy group in Washington DC. The FDA had estimated that it would need more than \$183 million in extra funding in 2012 to implement the food-safety act.

The spending bill holds both victories and disappointments for NOAA. The agency’s budget will increase, with an allocation of \$924 million to continue funding its Joint Polar Satellite System — a programme with a history of budget problems, but one that politicians are reluctant to attack because its data are crucial for forecasting severe weather. Not so fortunate was an effort to combine a series of government functions in a national climate service, a stated goal of Jane Lubchenco, who was appointed as NOAA director by Obama in 2009. With climate science a favourite target of Republican ire, the proposed service was turned down.

Another Obama appointee who is feeling the wrath of the right is John ▶

**“Compared to where we could have been, we’re in a good place for 2012.”**

➔ **NATURE.COM**  
For more on the US budget crisis, see:  
[go.nature.com/eg2noo](http://go.nature.com/eg2noo)

► Holdren, director of the White House Office of Science and Technology Policy (OSTP). Noting the OSTP's defiance of a 2011 congressional rule that bans the office from collaborating with China (see *Nature* 478, 294–295; 2011), the House appropriations committee hacked away nearly one-third of the office's relatively small budget, leaving Holdren with \$4.5 million to run his agency.

Neal Lane, a senior fellow at the James A. Baker III Institute for Public Policy at Rice University in Houston, Texas, and a one-time science adviser to former US president Bill Clinton, calls this direct attack by the appropriations committee on a part of Obama's administration an unusual move that is “very counterproductive” for science policy, and will have damaging consequences for the OSTP and US-based research as whole. “It's the only place in the federal government where anybody is paying attention to the overall health of science and technology in the country,” says Lane.

The action is one of many indications that the 2013 budget process, which will unfold during a presidential-election year, is likely to be hijacked and delayed by political manoeuvring. As *Nature* went to press, a bipartisan ‘super-committee’ of legislators tasked with finding a way to reduce the federal deficit by \$1.2 trillion

## A WORK IN PROGRESS

A spending bill enacted on 17 November includes 2012 budget allocations (US\$ millions) for several key federal science agencies. Still to come are final numbers for the National Institutes of Health, the Department of Energy and the Environmental Protection Agency, among others.

Agency	2010 actual	2011 estimated	2012 requested	2012 enacted
NASA	18,725	18,448	18,724	17,800
National Science Foundation	6,873	6,860	7,768	7,033
National Oceanic and Atmospheric Administration	4,853	4,588	5,498	4,894
Food and Drug Administration	2,362	2,447	2,744	2,497
National Institute of Standards and Technology	863	750	1,004	751

Sources: White House Office of Management and Budget; American Association for the Advancement of Science; FDA; US Senate Committee on Appropriations.

seemed all but certain to miss its 23 November deadline. This would trigger steep across-the-board cuts for all branches of the federal government, including all science-related agencies, beginning in January 2013 — unless Congress intervenes before then to overturn its own rules (see *Nature* 476, 133–134; 2011).

In the short term, legislators will move on to the next phase of their 2012 budget negotiations. By 16 December they must agree on a budget for those branches of the government not addressed in last week's spending bill. Among those are the Department of Energy

and the Environmental Protection Agency, frequent victims of Republican attack that were marked for significant cuts in 2011. Also to be determined is the allocation for the National Institutes of Health, the federal government's largest research enterprise at \$30 billion. The agency is waiting to learn whether its ambitious bid to launch a centre for translational science will survive this year's budget battles (see *Nature* 477, 141–142; 2011). ■

Additional reporting by Susan Young and Meredith Wadman.

## PARTICLE PHYSICS

# Higgs hunt enters endgame

*Large Hadron Collider could soon deliver a clear verdict on missing boson.*

BY GEOFF BRUMFIEL IN GENEVA, SWITZERLAND

Bill Murray is a man with secrets. Along with a handful of other scientists based at CERN, Europe's particle-physics facility near Geneva, Switzerland, Murray is one of the few researchers with access to the latest data on the Higgs boson — the most sought-after particle in physics.

Looking at his laptop, he traces a thin black line that wiggles across a shaded area at the centre of a graph. This is the fruit of his summer's labours. “It's interesting, actually, looking at this again,” he muses. A tantalizing pause. “But no, I can't say...”

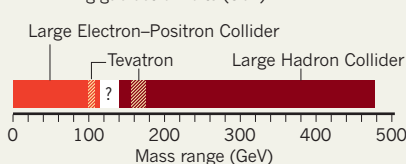
Despite Murray's coyness, there are few places left for the Higgs to hide. Billed as the particle that helps to confer mass on other matter, and the final missing piece in the ‘standard model’ of particle physics, the Higgs would be a huge prize for CERN's Large Hadron Collider (LHC), the world's most powerful particle accelerator. But so far, the two massive detectors there — ATLAS, where Murray works, and the Compact

Muon Solenoid (CMS) — have not seen any convincing signals of the elusive particle.

At a conference in Paris on 18 November, teams from ATLAS and the CMS experiments presented a combined analysis that wipes out a wide swathe of potential masses for the Higgs particle. Gone is the entire mass range from 141 to 476 gigaelectronvolts (GeV; energy and mass are interchangeable in particle physics). Together with earlier results from the 1990s, the analysis leaves a relatively narrow window of just 114–141 GeV in which the Higgs could lurk (see ‘Cornering the Higgs’).

## CORNERING THE HIGGS

Data from various particle accelerators restrict the mass of the Higgs particle, if it exists, to the range 114–141 gigaelectronvolts (GeV).



Analysis of the very latest data from this autumn — which Murray isn't yet ready to share — will scour the range that remains. If it turns out to be empty, physicists may have to accept that the particle simply isn't there. Working around the clock, the detector teams hope to have this larger data set analysed before the end of December. “We'll know the outcome within weeks,” says Guido Tonelli, spokesman for the CMS detector.

## WAITING FOR GOD

The quest for the Higgs boson, often called the ‘God particle’ after the title of a 1993 book by Nobel prizewinner Leon Lederman, is the public face of science at the LHC. Most high-energy physicists wince at the deistic designation, but they hold a near-religious devotion to the boson. Contrary to the popular view, their belief has less to do with mass than with fundamental forces.

Four fundamental forces are at work in nature: gravity, the strong nuclear force, the weak nuclear force and electromagnetism. Since the mid-1960s, physicists have strongly

## DO YOU BELIEVE?

Nature asked leading theoretical physicists whether they thought the Large Hadron Collider would find the Higgs particle predicted by the standard model of particle physics.



Tom Kibble, Imperial College London; Steven Weinberg, University of Texas at Austin; Sheldon Glashow and Lisa Randall, Harvard University, Cambridge, Massachusetts; Frank Wilczek, Massachusetts Institute of Technology, Cambridge; John Ellis, CERN, Geneva, Switzerland; David Gross, Kavli Institute for Theoretical Physics, Santa Barbara, California.

suspected that the weak and electromagnetic forces are actually different aspects of a single 'electroweak' force. This is partly because the photon, the force-carrying particle of electromagnetism, is highly similar to the force-carrying particles of the weak force — the W and Z bosons. Moreover, a single electroweak theory successfully predicts the interactions of fundamental particles.

There is one problem, however: the W and Z bosons are extremely heavy, nearly 100 GeV, whereas the photon is massless. To explain the difference, a number of physicists (including Peter Higgs in 1964) proposed a new field and particle. The eponymous Higgs mechanism would interact with the W and Z bosons, giving them mass, but would ignore the photon, allowing it to remain massless. Relatively straightforward tweaks to the Higgs machinery allow it to endow other particles, such as quarks, with their observed masses as well.

"The Higgs now sought at CERN is expected on the basis of the simplest picture" for the electroweak theory, says Steven Weinberg, a theorist who won a Nobel prize in 1979 for his work unifying electromagnetism and the weak force. "But there are other possibilities," he adds, reckoning the odds that the LHC's detectors will find the Higgs at 50/50 (see 'Do you believe?').

If there is no Higgs, then what? Gian Giudice, a theorist at CERN, recently published work suggesting that giant clusters of W bosons might serve the same purpose, but even he admits that "it would be a great surprise if it were true". Other models without the Higgs

boson invoke extra dimensions of space, but they are not yet sufficiently developed to guide experiments.

Perhaps the most

likely alternative is that the Higgs is not a single particle, but rather a class of particles, which together do the job of unifying the two forces. Such a concept might appeal theoretically if a single Higgs is not found, but it would be a major headache for experimentalists to check. Theorists believe that the conventional Higgs boson would leave only a subtle mark on the detectors as it decays into W and Z bosons, high-energy photons and other particles. If there were two Higgs-like particles instead of one, the signal of each would be weaker still, says Murray. "It starts to get quite messy to do the analysis," he says.

The answer to the Higgs question lies in the data now being crunched at CERN and other academic-computing centres around the world. The first 70 trillion or so collisions turned up intriguing Higgs-like decays in the ATLAS and CMS experiments, hinting at a particle of around 140 GeV (see *Nature* 475, 434; 2011). But the second batch of collisions showed nothing. If the collisions now being analysed show further evidence of Higgs decays, then the teams on the two experiments are likely to announce that they have found a tentative signal, to be firmed up in 2012. If not, the search will probably continue until the LHC is shut down for an upgrade at the end of next year.

Even if that continued search shows no evidence for a Higgs or anything else, the LHC will push on. Without a unified electroweak force, the standard model is unable to predict how certain particles and forces interact inside the collider, says Matthew Strassler, a theorist at Rutgers University in Piscataway, New Jersey. The LHC will gather data on exactly those processes, and that information could potentially be used to find a way in which electromagnetism and the weak force fit together. That process, Strassler adds, is likely to take many years. ■

## NATURE.COM

For video interviews with Higgs hunters at CERN, visit: [go.nature.com/ezpnaj](http://go.nature.com/ezpnaj)





The production of explosives such as nitrocellulose, used to fire shells, could be aided by synthetic biology.

## ETHICS

# Bioengineers debate use of military money

*US Department of Defense's call for greener ways to make explosives worries synthetic biologists.*

BY ERIKA CHECK HAYDEN

By transforming microbes into miniature chemical factories, synthetic biology could lead to cheaper medicines, greener fuels and, to the distress of some in the field, cleaner weapons. Last month, the US military asked synthetic biologists to look for more environmentally friendly ways to manufacture explosives, prompting a round of soul-searching about the field's dependence on defence dollars and whether some applications should be off-limits.

A 'statement of need' from the Strategic Environmental Research and Development Program (a partnership between the US Department of Defense, Department of Energy and the Environmental Protection Agency) calls for research on adapting biological systems so that they can be used in place of the current chemical methods for making explosives, which can involve toxic compounds such as heavy metals.

But manufacturing munitions is anathema to some scientists. "I can't look my kids in the eyes and tell them my ideas are being tossed around by generals. I'd rather they were thrown around by doctors or global-health researchers," says Eric Klavins, a synthetic biologist at

the University of Washington in Seattle.

The 'green explosives' project also highlights a broader debate. "We need to figure out what the right relationship is between the worlds of defence and synthetic biology," said Drew Endy, a bioengineer at Stanford University in California, at a defence-science conference in Las Vegas last week convened by the defence department's Defense Threat Reduction Agency in Fort Belvoir, Virginia.

Synthetic biologists have struggled to get much support from traditional biology funders, such as the National Institutes of Health, because they tend to regard the field as an engineering discipline, says bioengineer Pam Silver of Harvard Medical School in Boston, Massachusetts. They have therefore relied heavily on funding from the defence and energy departments, as well as the National Science Foundation.

Investments made by the Office of Naval Research during the 1990s, for instance, laid the foundation for the field, and the Defense Advanced Research Projects Agency (DARPA) in Arlington, Virginia, is set to award a series of contracts under its new synthetic-biology

programme, called Living Foundries.

Few of the military investments to date have been targeted at offensive applications, however. Jim Collins of Boston University in Massachusetts, for example, has been awarded US\$1.5 million a year for 5 years by the Office of Naval Research to develop what he calls "microbiorobots" — reprogrammed bacteria that are able to sense materials in the environment. These robots might eventually be used to detect underwater mines, for example, but Collins has a simpler goal: to demonstrate that living systems can be melded with electronic devices. Linda Chrisey, who manages the programme that awarded Collins's grant, points out that the Geneva Conventions forbid signatories from funding work that could facilitate the development of biological weapons.

But some researchers are uncomfortable receiving military funding — including the steering committee of the Critical Assessment of Genetically Engineered Networks, which aims to stimulate the field through a series of competitions and is sponsored by the W. M. Keck Foundation in Los Angeles, California. The committee decided not to apply for DARPA funding this year after committee member Klavins objected. "I understand that everything has a dual use. But who funds the ideas often determines which uses come first," says Klavins.

Roger Brent, of the Fred Hutchinson Cancer Research Center in Seattle, argues that although previous military-funded programmes in synthetic biology may not have had direct offensive uses, this is not the case for the green explosives call. He says that the field needs to decide whether it is prepared to follow other academic engineering disciplines, such as robotics, which have become heavily reliant on military funding and are involved in projects that have been used in warfare. "This could provide a good opportunity for the field to debate within itself whether this is a good direction to go in," says Brent.

And debate they will. Andrew Ellington of the University of Texas at Austin, who advises the Defense Intelligence Agency and holds grants from DARPA and the Office for Naval Research, says that the idea that scientists should not work with defence funding relies on a "1960s paranoid view of the military". He argues that if the United States is going to make weapons, it might as well make them using less-toxic processes. "You can have someone die because of the way we currently prepare munitions," he says, "or you can put in proposals to try to make it easier and safer and greener to make munitions."

Ellington adds that some of his projects — such as developing environmental sensors to detect mines — could even save lives. "I'm proud of the work I do. I'd like it if something we did someday kept somebody from getting blown up." ■

J. SUNDERMAN/US NAVY

## REGENERATIVE MEDICINE

# Stem-cell pioneer bows out

*Geron halts first-of-its-kind clinical trial for spinal therapy.*

BY MONYA BAKER

**T**he first company to test a human embryonic stem-cell product in patients has become the first big player to bail out of the field. Last week's move shook investor confidence and raised questions about whether the company had overreached itself, as well as underscoring just how difficult novel therapies are to develop.

Geron, based in Menlo Park, California, announced on 14 November that it would cease work on its stem-cell-therapy programmes to focus on its anti-cancer portfolio. "It's kind of a heartbreak," says Melissa Carpenter, principal of Carpenter Group Consulting in Seattle, Washington, who served as Geron's director of stem-cell biology in the company's early days. "It is truly unfortunate for the field that the first possible product didn't get to go out of the gate."

Human embryonic stem (ES) cells have the potential to turn into any of the body's cell types, and so could replace defective tissue in myriad diseases. Geron chose to pursue a treatment for spinal-cord injury — an ambitious goal, not least because spinal damage involves many cell types. Some in the field speculate that Geron went forward with the programme, at least in part, because the neural-cell precursors it was testing are relatively easy to derive from human ES cells, and because dramatic results in animal studies impressed investors when the company needed funding. However, human trials are often the only way to test such unprecedented therapies.

Only four of a planned eight patients in Geron's phase I trial have received injections of specialized cells derived from human ES cells. John Scarlett, who became Geron's chief executive in September (see 'Rough ride'), says that the company will continue to monitor enrolled patients, but will not recruit more. None of the four patients receiving stem-cell injections suffered serious adverse events, but there were also no hints that the therapy was working (although phase I trials are not designed to test for efficacy).

The price of Geron's stock fell by more than 30% at the news, from US\$2.28 to \$1.50 per share as *Nature* went to press. But the company estimates that discontinuing stem-cell research will save \$25 million a year, allowing

it to conduct half a dozen phase II clinical trials of its two cancer products in the next two years without raising additional funds. Cell-therapy studies would have taken longer and cost more.

Advanced Cell Technology (ACT) of Santa Monica, California, is now the only company conducting regulator-approved clinical trials involving human ES cells; these aim to treat degenerative eye diseases using specialized retinal cells.

Investors and patients are eager for reassurance that human ES cells have commercial and therapeutic potential, says Robert Lanza, chief scientific officer of ACT. "The field at this early point desperately needed a big success," he says. "It certainly puts a lot of pressure on us to deliver."

Geron's decision comes shortly after a ruling that products and processes involving human ES cells are not patentable in Europe (see *Nature* <http://dx.doi.org/10.1038/news.2011.597>; 2011), but analyst Reni Benjamin of investment bank Rodman & Renshaw in New York believes that Geron's decision was unrelated. It was probably already in the works when long-term chief executive Thomas Okarma left the company abruptly in February, and reflects a change in business strategy rather than a verdict on cell therapies in general, he says. "Just because Geron is out, it doesn't mean that other trials will slow down."

Geron had invested heavily to bring human ES cells to clinical trials. It funded the studies leading to the cells' derivation in 1998 (J. A. Thomson *et al.* *Science* **282**, 1145–1147; 1998) and burned through cash while devising ways to manufacture specialized cells from the stem cells, as well as running extensive animal tests to show that the cells were safe enough to use in humans. These early efforts paved the way for others trying to move stem cells into the clinic. "It's exponentially easier," says Lanza. "We know exactly what [regulators] want."

There may be other benefits for the field. Geron controls extensive intellectual property relevant to human ES-cell therapy, says Ken Taymor, a stem-cell patent expert at the University of California, Berkeley, and it may now be more willing to license this portfolio to help others pursuing such therapies.

Asked what other companies can learn from Geron's decision, Michael West, who ran Geron from 1990 to 1998 and is currently chief executive of biotech company BioTime of Alameda, California, simply suggests: "Don't be the first one out the door. The first one out the door gets all the arrows in his back." ■

## ROUGH RIDE

Geron led the field in bringing stem-cell therapies to clinical trial, but paid the price of being first.



**NOVEMBER 1990** Geron founded by Michael West.

**NOVEMBER 1998** Human embryonic stem cells (pictured below) derived and cultured for the first time, funded by Geron.

**MAY 2008** Despite a 21,000-page submission from Geron, the US Food and Drug Administration decides that its clinical trial of human embryonic stem cells should not proceed.

**JANUARY 2009** Clinical trial approved, just three days after US President Barack Obama's inauguration.

**AUGUST 2009** Clinical trial put on hold again.

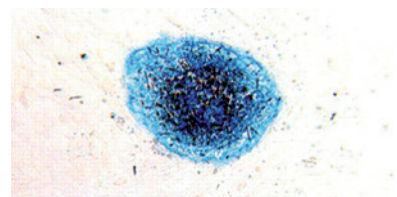
**JULY 2010** Hold lifted, clinical trial proceeds.

**OCTOBER 2010** Geron announces first dose given to patient in clinical trial.

**FEBRUARY 2011** Long-term chief executive Tom Okarma resigns.

**SEPTEMBER 2011** Geron appoints John Scarlett as chief executive.

**NOVEMBER 2011** Geron quits stem-cell research and development.



REDUX/EVINE

SCIENCE/AAAS





# THE MARS OBSERVER

**How the reclusive Mike Malin changed the way that scientists view Mars.**

**BY ERIC HAND**



It is sometimes said that Mike Malin knows Mars better than anyone else on Earth. A more verifiable statement is that Malin has seen more of Mars than anyone on Earth. His company, Malin Space Science Systems (MSSS) in San Diego, California, has designed cameras for every one of NASA's Mars-orbiting missions since Viking in 1975. Later this week, Malin will see the start of his ninth mission to the red

SANDY HUFFAKER/GETTY IMAGES, FOR NATURE

planet, when a launch window opens on 25 November for Curiosity, the US\$2.5-billion NASA rover that is carrying three of his camera systems.

Malin's devices are the eyes of the rover, the most costly and complicated Mars mission in a generation. The pictures taken by his cameras will help engineers to steer the machine; they will also be central to the scientific aim of the project: to determine whether Mars had suitable conditions for life billions of years ago.

When the images start streaming back from Curiosity, some nine months after launch, the best shots are likely to end up in Malin's hall of fame, the library inside his company's two-storey office building. A sanctuary in which the reclusive scientist can work alone, the library is strewn with copies of *Aviation Week & Space Technology*, which Malin says he has subscribed to since he was 14. In racks along the walls are poster-sized images of Mars taken from orbit. They display the wildly variegated terrain that has challenged scientists' understanding of the planet: scorched plains, dead volcanoes, mountainous dunes, chiselled canyons and the massive holes in the ground left by asteroid strikes. Everywhere he looks, Malin sees a land carved by wind, water and time. "I like these big prints," he says. "You see much more in them."

At 61, Malin no longer has the physical stamina for the geological field trips that took him to the ends of the Earth to help understand



analogous Martian landforms. But he still knows how to aggressively interrogate an image. He gets down on his hands and knees and, with a magnifying glass, puts his eyes inches above a Mars image he has laid flat. “What you can do with pictures is pretty limited,” he says. “But it’s mostly limited by your imagination.”

Prickly and driven, Malin is a unique force in the world of planetary science. Almost without exception, the instruments that NASA sends to other planets come from big government research centres and leading universities. But Malin’s company of 30 people has managed to corner the market both in the cameras that get sent to Mars and the discoveries that they provide. “Here’s this little company that’s been trusted to do it,” says Phil Christensen at Arizona State University in Tempe. “It’s amazing.”

Yet Malin feels that he struggles to gain respect — and to get his cameras onto missions. And he complains that other researchers have not sufficiently adopted his vision of Mars — as a planet with layers of sedimentary rock created and sculpted by water and wind. Yet the leaders of the Curiosity mission chose its landing spot, the Gale crater, precisely because it will allow them to study the sedimentary patterns there. Anybody but Malin would see that as a vindication of his work.

## THE WHOLE PICTURE

Malin’s cameras on Curiosity are sometimes overlooked amid the scientific gadgetry packed onto the 900-kilogram rover. One of those instruments is a laser that can vaporize distant rocks and, in the flash of light, look for tell-tale chemistry. Another will prepare thin samples for X-ray diffraction, a challenge even in a laboratory on Earth. Perhaps the most important instrument is the Sample Analysis at Mars, which will ingest rock samples and tease out their molecular make-up using gas chromatography and mass spectrometry in the hope of spotting an organic molecule — a sign that Mars once had a habitable environment.

The cameras are slightly more prosaic. They are much like modern digital cameras, but built to withstand the rigours of an interplanetary journey and years of dust storms and frigid temperatures on Mars’ surface. The largest of Malin’s systems are the Mastcams: two cameras on the rover’s neck that stand 2 metres off the ground and will survey the terrain. A second system, a microscopic imager fixed to the end of a robotic arm, will have sufficient resolution to see, for the first time, grains of silt. And a third, attached to the underside of the rover’s chassis, will monitor Curiosity’s descent and landing.

All these devices come from a man who once wanted to be a trombonist. Born in 1950 to a shoe-store manager and a secretary, Malin spent a large part of his teenage years obsessively practising the instrument. The extraordinary effort, he says, was triggered by a music teacher who suggested that he didn’t have the chops for it. He eventually became good enough to play with the Los Angeles Philharmonic and considered attending a conservatory. “If you ever want Mike to do something, tell him he can’t,” says Christensen. “He very much has an ‘I’ll show you’ personality.”

Nevertheless, Malin ended up choosing space science over the arts. He did his graduate work in the early 1970s at the California Institute of Technology (Caltech) in Pasadena, and was mentored by Bob Sharp, considered by many to be the father of planetary geology. Afterwards, Malin had a stint at NASA’s Jet Propulsion Laboratory at Caltech, a place with which he continues to have a love-hate relationship. He admires the lab’s planetary missions and its starry-eyed romanticism for the cosmos. But he disliked the bureaucracy, and often butted heads with managers. “I have a personality that chafes at large institutions,” he says.

Malin ran up against similar constraints at Arizona State University, which he joined in 1979. He thrived as a researcher, making a name through his geological studies in Antarctica — an analogue for Mars. But he grew to hate the committees and collaborations that are a part of so many university jobs. Malin is not necessarily antisocial, but he does do his best work alone. Christensen recalls a time when Malin invited him over for dinner, but the only furniture in Malin’s bachelor pad was an easy chair and a

television. They ate from trays in front of the television, Malin sitting in the chair and Christensen on the floor. “It never occurred to him to have more than one chair in his living room,” he says. And even though Malin now lives alone in the up-market coastal community of La Jolla, California, he doesn’t bother going to the beach.

In the mid-1980s, NASA began talking about a flagship Mars mission, the first since Viking, to be called Mars Observer. The many instruments on the orbiter — radiometers, spectrometers and laser altimeters — were expected to revolutionize the study of the planet’s surface, atmosphere and magnetic fields. Strangely, Malin’s call for an optical imaging camera drew little support. Few, he says, saw a need for mapping at a resolution sharper than the average of 50 metres per pixel that the Viking orbiters achieved. “There was this idea that we had mapped Mars with Viking, and there was no need to map it further,” he says.

But based on his fieldwork on Earth, Malin was convinced that greater visual detail would reveal the processes that played a part in shaping the surface. It was only after a last-minute intervention from NASA headquarters that Malin was given a contract to build and operate the Mars Orbiter Camera (MOC).

But where was he actually going to build it? The answer came in 1987, when Malin won a MacArthur ‘genius’ award and used the \$250,000 prize as seed money to found MSSS. There he developed a camera design for Mars Orbiter whose innovations would be copied on many subsequent missions.

Up to that point, Viking and other planetary missions had used framing cameras, which worked in the conventional way: a shutter opened and allowed light to hit an array of sensors. But Malin designed what is known as a push-broom camera. It had a permanently open aperture that funnelled light to a single line of charge-coupled devices (CCDs), at the time a new technology.

The continuous movement of the orbiting spacecraft provided the second dimension. Each time that data were read out from the line of CCDs, the spacecraft had moved forward a bit; an image was thus built

**“WHAT YOU CAN DO WITH PICTURES IS PRETTY LIMITED, BUT IT IS MOSTLY LIMITED BY YOUR IMAGINATION.”**

up over time, one line at a time. This design eliminated the need for any moving parts, such as shutters, in the camera — always a liability on a space mission — and allowed resolutions as good as 1.5 metres per pixel.

Before the camera could prove itself, though, the mission came to an abrupt end. Just a few days before Mars Observer was supposed to enter Martian orbit in August 1993, a fuel line ruptured and the craft flew off into space. Malin suddenly had to lay off half the employees in his fledgling business. But he soon earned a second chance. A replica of the MOC became the centrepiece of the 1996 Mars Global Surveyor mission, which entered Mars orbit in 1998. And as he prepared for the deluge of data that would come from the camera, the famously individualistic Malin realized that he would need a scientific sidekick.

He found one in Ken Edgett, who was just finishing his PhD under Christensen at Arizona State University. Today, Edgett works in an office near Malin’s second-floor library. A large, laid-back man, he slouches in his office chair with his shirt untucked. At the mention of a particular Mars image, he leans into his workstation and types in an MOC number: FHA-1858. “I know this one by heart,” he says.

That image changed everything in Mars science, Edgett says. It

➔ **NATURE.COM**

For more on the new Mars mission, see:

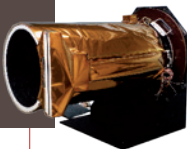
[go.nature.com/bclbtl](http://go.nature.com/bclbtl)

# FOCUS ON MARS

Michael Malin and his company, Malin Space Science Systems (MSSS), have built cameras for nearly all missions to Mars since the Viking programme in the 1970s.

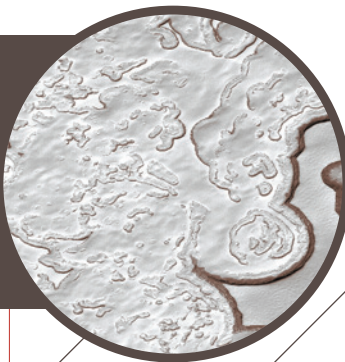
## NOVEMBER 1996

The Mars Global Surveyor spacecraft takes off with the Mars Orbiter Camera (MOC), a near clone of the imager on the failed Mars Observer. MOC operated from 1997 to 2006, taking more than 240,000 images.



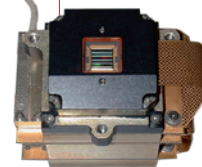
## FEBRUARY 2000

A MOC image of the south polar region reveals pits and other signs of erosion in the frozen carbon dioxide ice cap. The Swiss-cheese pattern indicates that Mars is undergoing climate change.



## APRIL 2001

Mars Odyssey launches, carrying the MSSS-built visible imager, part of a dual camera that captures both thermal and visible wavelengths.



1996

1997

1998

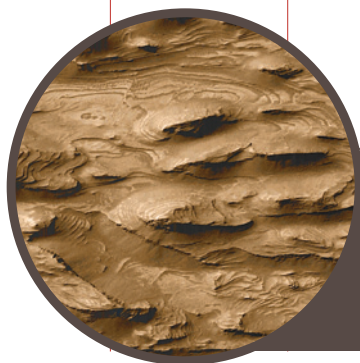
1999

2000

2001

2002

2003

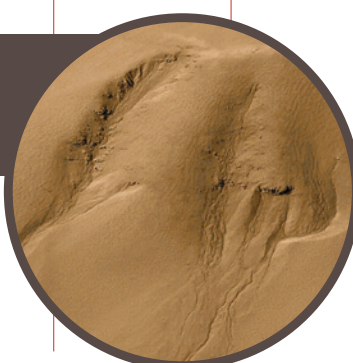


## MARCH 1999

The MOC discovers layered deposits at Candor Chasma, suggesting that sedimentary processes were important in shaping the Martian surface.

## SEPTEMBER 1999

The MOC discovers gullies on the wall of a crater, leading to the controversial suggestion that liquid water may still be active on Mars.



showed a region, called Candor Chasma, that had been carved into notched canyons and ziggurat-like steps<sup>1</sup> (see 'Focus on Mars'). The terracing had an enormous implication: these were once layers of sediment, probably laid down by water. And to get so many layers, water must have been at work for millions of years. It was the first solid evidence of a water-sculpted Mars.

The picture soon ended up on the cover of *Science*. It was one of four such covers for Malin and the MOC team, highlighting papers that touched on everything from the possibility of near-surface water flowing in large gullies today<sup>2</sup>, to the ever-changing Swiss cheese of frozen carbon dioxide at the poles that indicates ongoing climate change<sup>3</sup>. The MOC images also revealed the substantial sedimentary layering within a mysterious mound at the centre of the 154-kilometre-wide Gale crater — evidence that would eventually lead researchers to choose that site as the target for the Curiosity rover.

All those discoveries vindicated Malin, and his long push to get something like the MOC to fly. "It really changed our views of the planet," says Jim Bell, a geologist at Arizona State University who is the principal investigator for the cameras on Spirit and Opportunity, small rovers that landed on Mars in 2004. "It set the tone and the goals for the Mars programme that we're still pursuing today."

Not that Malin sought the limelight or found comfort in the recognition. He had to be dragged to press conferences, and cajoled into releasing data and writing up papers. He also demanded an unusual amount of control. With the MOC, Malin and Edgett personally chose the targets for nearly half of the 243,227 images it took. Because the MOC swivelled independently, the two researchers did not have to worry about any other instrument on the Mars Global Surveyor spacecraft. But on Curiosity, Malin's mast cameras will serve as the eyes for the other instruments. He will have to balance pursuing his own scientific goals with doing the reconnaissance needed to guide the sampling by the other instruments.

Some are worried that Malin and Edgett will find it hard to defer to the needs of others. "These are guys who like to do it their way," says Christensen. "They're not necessarily team players."

The members of Curiosity's science team will need to work together

if they are to crack the mystery at the centre of the Gale crater: a 5-kilometre-tall mountain of stacked sediments that reaches above the crater's rim. At some point, a massive amount of material must have filled the crater and buried it completely. Then erosive forces must have stripped most of the sediment away, leaving only the mound behind.

Curiosity is supposed to find out how sediment got into the hole, and how it got out. The working hypothesis is water and wind, acting over immense stretches of time. Researchers judge that the oldest sediments at the bottom were deposited 3.8 billion years ago, when Mars was thought to be warm and wet, and that the rock at the top of the mound represents a vastly different era some 200 million to 300 million years later, when Mars was growing colder and dryer<sup>4</sup>.

As it climbs upwards from the bottom of the mound, the rover will provide geologists with an unparalleled opportunity to study Martian history up close. By comparison, the Opportunity rover has traversed just 100 metres of strata in its nearly six years of exploration, according to John Grotzinger, Curiosity's project scientist at Caltech. "That is the gift of Gale," he says. "You know you're going to get a walk through time."

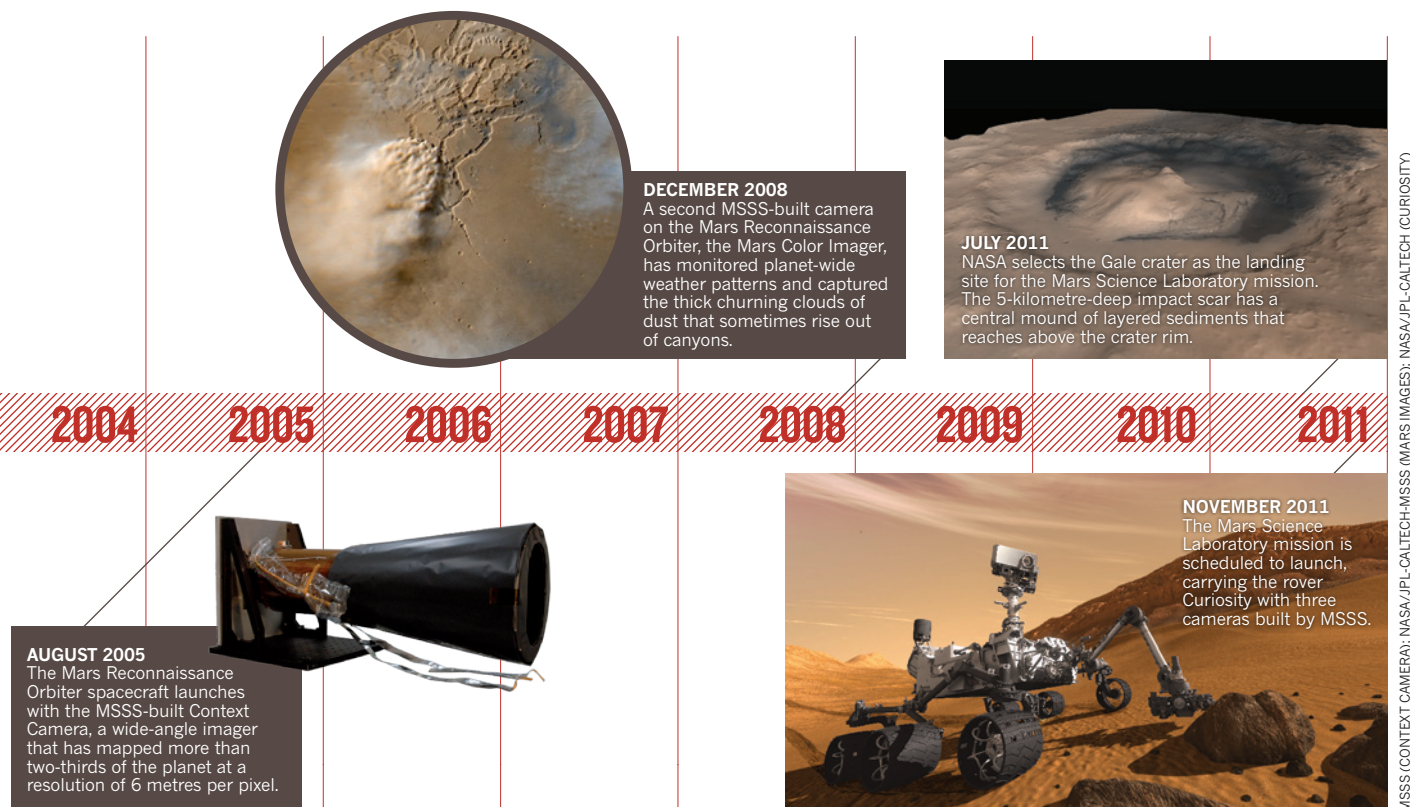
Although no one is willing to admit it, all on the team secretly hope that the rover, lasting a decade or more with its nuclear power source, could somehow make it to the top. A panoramic image taken from that vantage would surely be one for the ages — and yet it would be a bittersweet achievement for the man whose cameras would capture it. For although almost all the instruments on Curiosity are big steps up in capability from the sensors on Spirit and Opportunity, the Mastcams offer only an incremental improvement. That is not the way Malin had wanted the story to go.

## LOST FOCUS

Mike Ravine enters a storage room at the MSSS offices, flicks on the lights and considers two plastic boxes. He unlocks them and opens their lids — but does not touch the cameras inside, "The only time I've ever not delivered something was that," he says, stabbing his finger at the cameras. "It's disturbing to have been so close."

The advanced-projects manager at MSSS, Ravine is another employee





cut from the Malin cloth: clever, driven and a bit combustible. As well as working for MSSS, Ravine has launched an Internet start-up company, produced a movie and cultivated Hollywood contacts including the director James Cameron, whom he convinced to join the Curiosity camera team. They planned to develop twin 'zoom' cameras, which could switch rapidly between wide-angle panoramas and narrow, high-resolution close-ups. With high-definition and near-video capabilities on the cameras, Cameron would make three-dimensional movies for the public.

But with Curiosity running over budget in 2007, NASA demanded that the mission trim \$39 million from its instrument packages. The zoom cams had to be scrapped, even though Malin says that they were neither over budget nor behind schedule. Malin's team quickly designed two cheaper, fixed-focus cameras instead, one with a wide field of view, and one a narrow.

Ravine had one last chance to rescue the zoom cams when problems with the rover's motors forced NASA to delay the mission's launch from 2009 to 2011. He and Cameron flew to Washington to see NASA administrator Charles Bolden and convinced him to put the zooms back.

But by then, time was too short. There were problems with the alignment of the lenses and Malin's team could not fix them quickly enough. In March 2011, with most of the rover assembled and ready to go, NASA ordered the conventional cameras to be put on. Malin is still full of regret. He wonders how it would have gone had his team been able to work on the zoom cams all along. "They should never have been de-scoped," he says.

In February, just a few weeks before NASA shelved the zoom cameras for good, Malin had a heart attack. He'd already had one, two decades ago, and a stroke six years ago. But neither deterred him from long days and nights of work.

This time was different. Malin needed a quadruple bypass. As he recovered from surgery, his absence at the office was palpable. His employees — the closest thing Malin has to a family — wondered about their future.

Six months later, Malin is back, tethered to an iPhone that reminds him to take his medicine. Doctors have told him to curtail his workaholic

hours. But on this September day at the office, Malin is as intense as ever. He enters an empty, cavernous room about the size of a basketball court, and offers a glimpse of his vision for the future of Mars exploration.

## MARS ON EARTH

"As a kid I thought I'd be the first person on Mars," he says. "That ain't gonna happen." Instead, he wants to bring Mars to Earth, or at least to use his cameras to give people a feel for the richness of another world. Malin paces off a circle in his empty room, winking at an imaginary point in the centre to mimic the snapshots he would like his Mastcams to take on Mars. In this painstaking way, he would build a three-dimensional data set of a rock or other object of interest to scientists. Then, with the help of a virtual-reality helmet, Malin could recreate the view back in this room in San Diego. Geologists could peer in close, look behind a rock, or even explore it virtually using the rover's arm. And, ever the business man, Malin says that people might pay for the virtual trip to Gale crater. "Basically, you'd walk around — and you'd be on Mars!" he says.

Malin has told some of his colleagues about the idea. Although they've offered murmurs of support, he knows that it will be up to him to make things happen. It will certainly be tougher without the zoom cameras or the full support of an Oscar-winning director. But never count Malin out. It took him years to convince his colleagues of the importance of a sedimentary Mars — and now they're sending a multibillion-dollar rover to the largest mound of stacked sediment in the Solar System.

Standing in the vacant room, Malin surveys the space. But you know that, in his mind, it's already a Martian landscape. "It's not real to people because they don't see it," he says. "I'm going to make it real." ■

SEE EDITORIAL P.446

**Eric Hand** covers physical sciences for *Nature* in Washington DC.

1. Malin M. C. & Edgett K. S. *Science* **290**, 1927–1937 (2000).
2. Malin M. C. & Edgett K. S. *Science* **288**, 2330–2335 (2000).
3. Malin, M. C., Caplinger, M. A. & Davis, S. D. *Science* **294**, 2146–2148 (2001).
4. Thomson, B. J. *et al.* *Icarus* **214**, 413–432 (2011).





# Spring comes to ancient Egypt

BY JO MARCHANT

As the country struggles to refashion its government, archaeologists are looking warily towards the future.

In a secluded stretch of desert about 300 kilometres south of Cairo, hundreds of bodies lie buried in the sand. Wrapped in linen and rolled up in stiff mats made of sticks, they are little more than bones. But their ornate plaited hair styles and simple personal possessions help to reveal details about the individuals in each grave. The bodies date from around 3,300 years ago, when the Pharaoh Akhenaten renounced Egypt's traditional polytheistic religion and moved his capital to remote Amarna, to worship just one god: the Sun disc Aten.

The cemetery offers a window on a unique episode in Egyptian history, a revolution that some see as the birth of monotheism. Barry Kemp, an archaeologist at the University of Cambridge, UK, and director of the Amarna Project, has been working with his colleagues to excavate the skeletons, and says that they are starting to reveal "an alarming picture of a stressful life". Many Amarnans died young, with retarded growth and signs of multiple injuries. Some young men had marks where their shoulder blades had been pierced, perhaps as part of a brutal ritual.

**The Egyptian Museum in Cairo had to be guarded from looters during civil unrest in January.**

T. TODRAS-WHITEHILL/AP

Kemp can't say much more about the skeletons because he had to flee the site in January, putting his team on flights out of the country and walling up his storehouses as a present-day revolution sent the country into chaos (see 'Archaeology in turmoil'). Although the situation soon calmed — in fact, Amarna did not suffer a single episode of looting — Kemp has spent months waiting for permission to resume excavations. Other teams working in the country tell a similar story. "We've lost a year," says Frank Rühli, a palaeopathologist from the Centre for Evolutionary Medicine at the University of Zurich, Switzerland, who was scheduled to start work in February on human remains at the pyramids of Saqqara, near Cairo, and in the Valley of the Kings near Luxor.

The block on excavations has been the latest in a series of obstacles for archaeologists working in Egypt — the home of perhaps one-third of the world's antiquities, which reveal a vanished culture in unmatched detail (see 'New research in an ancient land').

Egyptian officials have said that their reluctance to allow work to restart stems from security concerns; they are now starting to grant permits for excavations. But a broader problem is that Egypt's Supreme Council of Antiquities (SCA), which coordinates all conservation and excavation activities in the country, has been mostly paralysed since the departure of its charismatic but controversial leader, Zahi Hawass. An ally of Egypt's deposed president, Hosni Mubarak, Hawass was forced to leave office in July. Since then, the agency has gained and lost three heads in quick succession, with the latest secretary-general, Mustafa Amin, appointed at the start of October.

The uncertainty dashed hopes of a swift return to normality for archaeological research, and unrest this week adds new concerns. "Everything is up in the air," said Kim Duistermaat, director of the Netherlands–Flemish Institute in Cairo, last month. As Egypt struggles to determine its future without Mubarak, archaeologists are wondering what their field might look like without Hawass.

## RISE OF THE PHAROAH

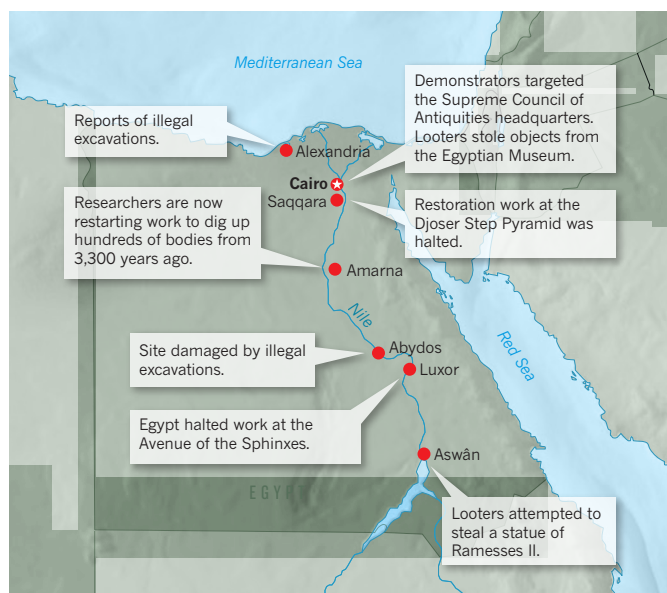
The antiquities service was set up in 1858 to stem a different kind of chaos: the loss of artefacts. Early Egyptologists were little more than treasure hunters, who carted off everything from jewellery to entire monuments. Now, the SCA conducts its own excavations and approves and supervises foreign archaeological missions, as well as conserving and managing the country's wealth of antiquities and archaeological sites.

The service was initially led by French scholars, and did not have an Egyptian head until the 1950s. After becoming secretary-general of the SCA in 2002, Hawass catapulted what had been a fairly anonymous position into the limelight. He mixed with celebrities from Diana, Princess of Wales, to US President Barack Obama; fronted big-budget television documentaries; and even starred in his own reality show, *Chasing Mummies*. The image of Hawass enthusiastically unearthing treasures in his Indiana Jones-style hat became a familiar sight, and it gave Egyptology its first Egyptian face.

Even as he raised his own profile, Hawass did the same for archaeology in Egypt. His efforts attracted tourists and raised millions of dollars from international touring exhibitions of Tutankhamun's treasures. He fought hard — some felt too hard — for repatriation of artefacts, and pushed for Egyptian teams to conduct high-profile science (see *Nature* 472, 404–406; 2011). He raised money for state-of-the-art facilities in Egypt, notably persuading National Geographic in Washington DC to donate a US\$3-million scanner to the SCA in return for filming a project to scan Tutankhamun and other royal mummies; US broadcaster the Discovery Channel built two ancient-DNA labs in Cairo and donated \$250,000 towards testing the mummies' DNA. Hawass also tackled corruption and supported projects to develop archaeological sites, including building a suite of museums and dealing with rising groundwater that is threatening to damage sites across the country, including Giza's famous pyramids.

But critics claim that Hawass had a darker side: that as the years went on, he exerted excessive

**► NATURE.COM**  
For more on Egyptian  
archaeology, visit:  
[go.nature.com/jachks](http://go.nature.com/jachks)



## Archaeology in turmoil

Researchers in Egypt are trying to resume work at excavations that were shut down during the revolution in January. Looting and illegal building have plagued many important sites.

control and sought mainly to boost his own fame at the expense of other researchers and of high-quality science. Under Hawass, they complain, archaeologists were prevented from announcing their own discoveries. "This focus on him was something that really bothered people," says Duistermaat. "Even for foreign missions, you had to wait, even for weeks, until Zahi would come down and 'excavate' it."

Many archaeologists working in Egypt are reluctant to speak about Hawass on the record out of fear that he could regain influence in the country. But in private, several researchers say that Hawass was intolerant of opposition and blocked excavation permits to those who published results or theories that clashed with his own. Megan Rowland of the University of Cambridge, who has just completed a master's of philosophy degree on the political significance of Egypt's antiquities during the revolution, says that researchers who crossed Hawass became targets of intense criticism or had their permits revoked.

"Egyptological research is subject to very heavy censorship," she argues. In media interviews over the years, Hawass has accused several well known archaeologists of smuggling, scientific fraud or other improprieties. One researcher targeted by Hawass was Joann Fletcher, an Egyptologist from the University of York, UK. In a 2003 television documentary she suggested that a particular mummy was Queen Nefertiti, wife of Akhenaten — a finding that Hawass says he did not vet, and which was at odds with his own ideas.

Hawass told the Australian television programme *60 Minutes*, "It is clear [Fletcher] made all this up because she wants to be famous." Fletcher was temporarily blocked from excavating in Egypt. She challenges Hawass's account and maintains that she did not break any rules.

Researchers also face restrictions when they seek to analyse artefacts. Despite Hawass's efforts, Egypt still has only limited capacity for sophisticated testing, such as carbon-14 dating and DNA analysis. But it is illegal to remove any archaeological artefact — even mud or pollen samples — from the country for analysis. Although some see this as an understandable response to the history of artefacts being illicitly exported, others complain that it is devastating for archaeological science. "This is what makes us look like fools at international conferences," says one researcher, based in Cairo, who does not wish to be named.

Just a year ago, it seemed impossible to imagine any change in this situation. The position of SCA secretary-general has traditionally been temporary, held for just two to three years. But Hawass had the support





Protesters filled Tahrir Square in Cairo, near the Egyptian Museum (left), during the revolution, and the antiquities council lost its leader, Zahi Hawass (right).

of Mubarak, who extended his appointment.

The revolution changed all that. Hawass's hold on power started to slip when he denied, incorrectly, that any objects were missing after Cairo's Egyptian Museum was looted on 28 January. It was further eroded when he underestimated the extent of looting at important sites, despite reports that it was severe, and repeatedly voiced support for Mubarak. When Mubarak fell, Hawass's days were numbered. After resigning and being reappointed in March, Hawass finally left office in July.

He has barely appeared in public since, and has been under investigation by the Office of the Attorney General for a range of alleged offenses including stealing artefacts and diverting money from a touring Tutankhamun exhibition to a private charity owned by Mubarak's wife, Suzanne. "It's laughable," says Salima Ikram, an Egyptologist at the American University in Cairo, who has worked in Egypt for 18 years. "Zahi would never steal antiquities."

Today, Egypt's most famous archaeologist can be found tucked away on the ninth floor of a faded apartment block in the Mohandessin district of Cairo. Forbidden to leave Egypt while the investigation is ongoing, Hawass spends his days writing books in this modest office, surrounded by trophies, medals and photos of himself with celebrities. When *Nature* visits, he is charming and full of energy, bouncing up from his desk every few minutes to locate objects that will illustrate a point: his sweat-stained hat; his handwritten manuscripts; and a tall pile of stuffed envelopes that he says will prove his innocence in the attorney-general's investigation.

Hawass denies having close ties to Mubarak and calls the charges against him "ridiculous and untrue". Almost all of them have been dismissed, and the rest will soon be resolved, he says. Regarding his leadership style and appearances on television, Hawass says that it was important for him to maintain a high profile "to Egyptianize Egyptian antiquities". He denies taking credit for others' discoveries, arguing that he was required to scrutinize all results before they were announced to the media, to prevent unscrupulous archaeologists from making false claims. "Many people announce wrong information to get money," he says. He acknowledges that people have been banned from working in Egypt, but says that such decisions were made by a 60-person committee of the SCA and the sanctions were imposed only when researchers did not have proper credentials or broke SCA rules, such as announcing findings without approval.

Rather than harming Egyptian science, Hawass says that he raised standards, cleaned up corruption and trained a new generation of researchers. Hawass sees his work — and his ability to extract money

from foreign television companies — as a high-profile success for Egyptian Egyptology. "I'm very proud of the results," he says, describing the paper reporting DNA analysis of Tutankhamun (Z. Hawass *et al.* *J. Am. Med. Assoc.* 303, 638–647; 2010) as "an incredible article". High-profile projects like that, he says, help to "raise the global interest in Egyptology".

But foreign researchers have criticized the studies, complaining that raw data were not shared, making it impossible for them to assess the quality of — let alone repeat — the work. Some complain that the research was carried out purely for television audiences, whereas less glamorous projects might have had greater scientific value.

## BACK TO BUSINESS

Love him or hate him, Hawass's departure has unnerved Egyptologists. Asked what they're hoping for from his successor, many researchers say that they want more open discussion of ideas, more sharing of data and collaborations between Egyptian and foreign teams.

But first, the SCA needs to get back on its feet. Researchers had hoped to resume work as soon as the security situation calmed. But the agency has been dogged by protests since the revolution, and Hawass's departure left it in chaos. None of his successors at the SCA has yet managed to last more than two months, and researchers say that progress has stalled.

"This is the first time in the course of five administrators I've lived through as an adult Egyptologist that it can't function," Ikram said last month. As researchers waited through the summer, permits were left unsigned and decisions unmade. When *Nature* visited in October, the agency's headquarters in Zamalek, Cairo, was a hive of inactivity, with dozens of men milling around its halls and the waiting room filled with bored employees watching the clock until it was time to go home.

"We've been sitting here for six months," said one, clearly frustrated.

Everyone now hopes that Amin, the SCA's latest secretary-general, can get things started again. He holds a PhD in Islamic antiquities, and was previously head of the SCA's Islamic and Coptic department. Researchers say it is too early to comment on his leadership style, but because he does not specialize in Egyptology, it seems unlikely that he will share Hawass's one-man approach — or front documentaries about the pharaohs.

"He'll need people beside him," says Atef Abu El-Dahab, the affable head of the SCA's Egyptian antiquities sector. "First of all, me."

LEFT: K. DESOUKI/AFP/GETTY IMAGES; RIGHT: XINHUA/EYEVINE

## BURIED TREASURE

*New research in an ancient land*

Despite problems that have plagued Egyptology in recent years, the country continues to attract foreign researchers because of its special attributes. “Egypt is one of the few countries that maintained its borders and culture for 3,000 years,” says Salima Ikram, an Egyptologist at the American University in Cairo. “It’s unique in its historic and cultural span.”

Key research areas include the origins of the country’s civilization, how Egyptian society stayed stable for so long and the influence of climate change. Chris Naunton, director of the Egypt Exploration Society (EES) in London, says that rather than studying “old fashioned” tombs and temples, the EES is increasingly

interested in “settlement archaeology” and how people used the land. One project, led by David Jeffreys, an archaeologist at University College London, is looking at how the path of the Nile river has shifted over time.

Techniques such as DNA analysis are encouraging archaeologists to focus on human remains, which can provide information about everything from hairstyles to the evolution of disease. Ikram is collaborating with David Lambert, a biologist at Griffith University in Brisbane, Australia, to extract DNA from mummies of ibises — sacred birds in ancient Egypt — to study rates of mutation over time.

The country has a wealth of texts that

can be cross-referenced with archaeological finds, and a dry climate that beautifully preserves organic materials. André Veldmeijer, assistant director of the Netherlands–Flemish Institute in Cairo, is studying leather trappings from a chariot, rediscovered last year in the Egyptian Museum in Cairo (right). The find includes harnesses, gauntlets and a bow case, complete with wear marks and an elaborate red-and-green design, still bright after 3,000 years. **J.M.**



Amin has some huge problems to address before even thinking about boosting the quality of research. His first priority is the security of Egypt’s sites and museums. Some looting is still going on, and the full extent of the losses isn’t known, says Tarek El Awady, director of the Egyptian Museum in Cairo. “We’re still waiting for the inventories,” he notes. But the most serious challenge is illegal building, with locals trying to claim archaeological land at several sites.

El Awady says the underlying issue is that local people don’t appreciate the importance of the country’s archaeological heritage. Rowland blames this alienation on the Mubarak regime’s “highly politicized approach to heritage management”. She argues that Hawass had absolute power and focused on foreign audiences, which left local people with no sense of ownership of their own antiquities.

But El Awady defends his former boss. “He played an important role in increasing people’s knowledge of Egyptian heritage,” he says. Still, he adds, the looting shows the importance of “building bridges between museums, sites and local communities”.

**LOST MILLIONS**

The second major problem facing Amin is funding. The SCA had a healthy income during Hawass’s tenure, but the coffers are now empty, despite the extra millions of dollars that should have come in from the travelling exhibitions. “We have no money,” confirms El-Dahab. He says that all conservation and excavation projects have been halted, and the agency is now borrowing millions of dollars from banks and the government just to pay salaries.

There is no shortage of conspiracy theories as to what might have happened to the cash, but El-Dahab says that it has gone to the many projects that Hawass championed, including the construction of 22 local museums, conservation and restoration work at important sites, and his efforts to deal with rising groundwater.

Hawass denies any impropriety and defends his record. “I spent 1 billion Egyptian pounds [US\$167 million] a year” in support of Egyptian archaeology, he says proudly. He adds that he had planned to bring in more funds through tourism and travelling exhibitions, and blames the political situation — which has drastically cut the number of foreign visitors — for the SCA’s financial crisis.

To make matters worse, many of the agency’s employees have been angrily protesting since the revolution for better pay and conditions, blockading SCA buildings and obstructing tourists. The agency has a huge staff — a spokesperson refused to even guess how many, but Egyptologists estimate that there are perhaps 40,000 permanent employees

and another 15,000 or so on temporary contracts. But the SCA doesn’t have the money to pay them, or enough work for them all to do. El Awady says that a large proportion of SCA staff should be let go: “We don’t need all these workers.”

However, it seems certain that there will not be large numbers of layoffs. The protesters forced out Amin’s three predecessors, and Amin will need to keep employees on his side. He is now negotiating with the government for the funds to provide them all with permanent contracts. Amin also announced in October that he will carry out a comprehensive inventory of all SCA-owned land, selling or leasing any areas declared free of monuments and artefacts in order to raise money.

He promises to revive restoration work at the pyramid of Djoser, Egypt’s oldest surviving stone building, and other major projects — if the government gives him the money. Meanwhile, permissions for foreign research are starting to come through. Kemp’s group finally returned to the field earlier this month.

Ultimately, however, the future of archaeology in Egypt depends not just on Amin, but on the outcome of Egypt’s first democratic elections in decades, scheduled to begin on 28 November. Researchers are wondering whether the new political regime will take a nationalistic approach that favours Egyptian researchers, or become more open to foreign researchers and international collaborations.

And there is one more move that the new government could make. Egypt is reliant on funds from the millions of tourists who come to see its antiquities each year, and although visitor numbers have picked up slightly since the revolution, they are still low. El-Dahab says that the number of tourists visiting the country in September 2011 was only one-quarter of what would normally be expected.

If there was one thing that Hawass was good at, it was bringing in tourists, keen to visit after watching his exploits on television, or marvelling at Tutankhamun’s travelling treasures. So it is not inconceivable that a new leader might yet invite the charismatic archaeologist back to the SCA.

Hawass has previously denied any interest in returning to his old job, but now seems to be repositioning himself. “I’m sorry to say it, but I’m the only one who can bring the tourists back,” he told *Nature*. So would he offer his services, if asked? “I will never come back unless there is a stable government,” he says. If the upcoming elections can deliver that, the man in the hat might yet rise again. ■ **SEE EDITORIAL P.445**

**Jo Marchant** is author of *Decoding the Heavens: Solving the Mystery of the World’s First Computer* (William Heinemann, 2008).



# COMMENT

**POLICY** African farms need a genetic revolution **p.471**



**BIOGRAPHY** The beautiful mind of screen goddess Hedy Lamarr **p.474**

**FICTION** Tiny is terrifying in Michael Crichton's last thriller **p.476**

**CORRESPONDENCE** Damming the Mekong could damage public health **p.478**

A. JAMAL/WFP



The World Food Programme encourages local, innovative food solutions, such as Wawa Mum, a fortified chickpea paste developed and produced in Pakistan.

## Preventing hunger: Sustainability not aid

Fifty years after its founding, UN World Food Programme head **Josette Sheeran** explains why the agency is now focusing on projects that help communities weather food crises.

**N**ever before has food been so abundant. Global agriculture produces an estimated 17% more calories per person than it did 30 years ago, despite a population increase of more than 50%. Yet each morning, nearly 1 billion people wake up hungry, most of them women and children. Although gains have been made in reducing the global population of those

undernourished — from about 26% in 1971 to 13% in 2008 — the current global economic crisis has pushed millions more into poverty and swelled the ranks of the hungry.

Fifty years ago, in December 1961, the General Assembly of the United Nations approved a resolution that established the World Food Programme (WFP). It was given the goal of saving lives and livelihoods

— bringing food from lands of plenty in the post-war years to those in need. Within a year of its founding, the WFP would undertake three major operations — supplying food for survivors of the disastrous 1962 earthquake in Iran, victims of a typhoon in Thailand and refugees returning to newly independent Algeria. Since then, the WFP has continued to respond to emergencies ►

► around the globe and is now recognized as the world's largest humanitarian organization. Last year alone, it brought food and nutrition support to more than 109 million hungry people in 75 countries.

Food aid has saved millions of lives, but it cannot, by itself, solve hunger. This is why, over the past few years, the WFP has been undergoing one of the most profound transformations in its history, as it moves from a food-aid agency to one that helps individuals, communities and nations to build sustainable food security. In this way its response is smarter, more targeted, and it is able to help the most vulnerable people prepare to withstand shocks — whether from natural disaster, economic crises or food price volatility.

Just as the WFP has evolved since its founding, so too has the environment in which it operates. The difficulties of ensuring access to nutritious food for the most vulnerable people have been compounded by a convergence of global trends: continued food price volatility; an increase in the number and intensity of weather-related disasters; and an escalation of ongoing conflicts.

## ROOT CAUSES

In 2007 and 2008, when global food and fuel prices skyrocketed and sparked food riots in 35 countries, more than 115 million people were added to the ranks of the hungry. Food prices are again surging on global markets. The World Bank estimates that as a result of these price rises, another 44 million people were pushed into extreme poverty between June 2010 and February 2011. It is those who are extremely poor and vulnerable who suffer the most — women and girls often have disproportionately less food during economic shocks. Families are forced to sacrifice tomorrow for today — eating income-producing livestock, putting schoolchildren to work and switching from more expensive, nutritious food to cheaper staples.

Climatic changes have increased both the frequency and the intensity of natural disasters. Floods and droughts that were once occasional have now become epic and more regular. The 2011 drought in the Horn of Africa was declared the worst in 60 years. The 2010 monsoon floods in Pakistan were the worst in the country's recorded history.

And escalating conflict and political instability have plunged millions around the world into food insecurity. Nowhere have we seen this play out more dramatically than in the Horn of Africa, where Somalis have endured two decades of civil war and two consecutive seasons of failed rains. Now, after their livestock and crops have died, they are faced with the terrible choice left to people without food: migrate or die.

Make no mistake; this is not the failure of aid — but the lack of access to aid.

Although we cannot prevent drought, we can prevent famine. There are hopeful signs that outside Somalia, where those in control have blocked humanitarian assistance, the drought's impact has been blunted by advance preparation and putting in place programmes that help the most vulnerable populations to better weather such crises.

Through the Managing Environmental Resources to Enable Transitions to More Sustainable Livelihoods (MERET) programme, the WFP has been supporting the Ethiopian government in sustainable land management and rain catchment, which has vastly increased food production and mitigated the impact of the drought. In the dry Karamoja region in Uganda, local communities have established a system of communal food stocks that are replenished at harvest time, enabling them to cope with periods of food insecurity.

The WFP is working with countries across the globe at the grass-roots level to develop — and scale up — innovative ideas and tools to transform the fight against hunger. In Cameroon, for example, where about 2.8 million people are food insecure and the lean season in the north of the country lasts an average of three to four months, every year can be a crisis for the most vulnerable people. To help break the boom-and-bust cycles of hunger,

*"Since its founding, the World Food Programme has worked itself out of a job in more than 50 countries."*

the WFP provides a one-time donation of 10 tonnes of cereal for each community granary and helps to train farmers in food-storage management and financial accounting. Community members can withdraw stocks from the granary during the lean season, and later replenish from their own crops during harvest, paying little interest. The steering committee for each granary uses the revenue collected from interest and sales of commodities to reconstitute stocks and ensure the village's access to affordable food all year round.

Working with food technologists, we are deploying products such as Wawa Mum. This highly fortified chickpea paste, developed and produced in Pakistan, requires no water or cooking, and contains essential micronutrients for young children, who suffer irreversible damage to their minds and bodies if they don't receive sufficient nutrition in the first 1,000 days of life — from conception until they are 24 months old.

The WFP is also using technology to reach the most vulnerable people and support local economies, through 'digital food' such as electronic vouchers delivered to mobile phones. In the Palestinian territories, for example, beneficiaries are able to use an electronic swipe card to purchase nutritious food at local markets. All the products in

the programme, such as milk, yogurt and cheese, are produced locally.

These WFP programmes are fundamentally different from food aid that is often brought from the outside. They enhance food security by ensuring that our responses are supporting local markets and farmers, and enabling residents to buy locally produced products that might otherwise be out of reach. Where the digital-food programme is in operation, local dairy farmers have increased production by 30%. Local shops have more customers and higher profits. And people can choose nutritious food for their families in a way that protects their dignity. Because of this innovative approach to promoting food security, the WFP is increasing cost-efficiency and enabling a better analysis of food consumption patterns in real time.

## THE TOOLS ARE AT HAND

As these examples show, ending hunger does not require a major scientific breakthrough such as those that have graced the pages of this journal. Preventing the loss of a generation of children due to malnutrition is, on one level, relatively simple: women and children need access to an adequate amount of nutritious food. For the first time in history we have the scientific knowledge, programmes, tools and policies to defeat hunger but we need global political will. It requires the commitment of national leaders to put in place the right policies and to declare that people will not starve under their watch.

Former Brazilian President Luiz Inácio Lula da Silva made it clear that fighting hunger would be his top priority, stating in his inaugural address that "if at the end of my term every Brazilian person has three meals per day, I will have fulfilled my life's mission". Since 2003, the *Fome Zero* (Zero Hunger) programme, which focuses on access to food for the poorest and support for small-scale and family farmers, has helped to halve the proportion of hungry people in Brazil. The country was a WFP beneficiary in the 1990s — now it is our ninth-largest donor.

Since its founding, the WFP has worked itself out of a job in more than 50 countries as nations take over programmes such as school feeding and other safety-net projects — a testament to what can be achieved when national leaders, the private sector and civil society come together and collectively declare "not on our watch". We may always need a WFP when unexpected disaster strikes. But 50 years from now it is my dream that no one will wake up hungry and every child will have sufficient nourishment to reach their potential. ■

**Josette Sheeran** is the executive director of the United Nations World Food Programme, 00148 Rome, Italy.  
e-mail: [wfpinfo@wfp.org](mailto:wfpinfo@wfp.org)





Planting crops that are drought resistant could enable farmers to use less water and fertilizer.

# Preventing hunger: Biotechnology is key

If African countries can't plant genetically modified crops to produce more and healthier food, vulnerable populations will be at risk, argues **Calestous Juma**.

**T**o survive the droughts, wars and other major causes of famine, Africa must embrace technologies that enable it to produce more, better food with less effort.

Indeed, without the advances in molecular biology and other scientific fields that occurred in the second half of the twentieth century, African nations would be much worse off than they are now. Without this Green Revolution, which enabled developing nations to import cheaper grains and grow high-yield seed varieties, analysts estimate that crop yields in developing countries would have been 23.5% lower and prices between 35% and 66% higher in 2000. Caloric intake would have dropped by up to 14.4%, and the proportion of malnourished children would have increased by nearly 8%. Put another way, the Green Revolution helped to raise the nutritional status of up to 42 million preschool children in developing countries (R. E. Evenson and D. Gollin *Science* **300**, 758–762; 2003).

These tools were a great help to African nations in the previous century, but they are not sufficient to help Africa's agriculture

survive what is coming: rising population and loss of productivity brought on by ecological disruptions such as environmental degradation and frequent droughts.

To weather these changes, African nations must be open to new biotechnology tools that allow farmers to grow crops that have even higher yields and a higher nutritional content, and which can withstand biological and physical stresses.

At present, only a few African countries are allowed to grow genetically modified (GM) crops, partly because of restrictive national biosafety policies that impose excessive regulatory barriers to the adoption of agricultural biotechnology. This must change.

For starters, African farmers need pest-resistant GM cotton, which is already being cultivated in South Africa and Burkina Faso. These crops do not raise food-safety concerns, but their higher yields bring more disposable income to farmers, who can use that money for food crops. More countries should be planting herbicide-tolerant maize (corn), now in use in South Africa and Egypt, which reduces the need for

weeding — a significant benefit for African farmers, mostly women, who spend nearly 2,000 hours a year weeding a hectare of land.

Future innovations could bring even more benefits to African countries. For example, the Bill & Melinda Gates Foundation of Seattle, Washington, is funding projects to produce crops that carry drought-tolerant genes.

Africa needs GM varieties of the black-eyed pea, a subspecies of the cowpea (*Vigna unguiculata*). The crop is often attacked by the insect *Maruca vitrata*, which causes US\$300 million in losses annually to small-scale farmers in Africa. Their only means of controlling the disease is using expensive pesticides, which cost Nigeria an estimated \$500 million a year. Their losses affect the world: the African continent currently produces nearly 5.2 million tonnes of the bean, accounting for nearly 70% of global output. Scientists at the Institute for Agricultural Research at Ahmadu Bello University in Zaria, Nigeria, have developed a GM variety containing insecticide genes from the bacterium *Bacillus thuringiensis* that can help to control the disease.

## DISEASE RESISTANCE

Africa would also benefit greatly from having other disease-resistant crops. Bananas, for example, are a staple in Uganda, where adults often eat three times their body weight in bananas annually. But the banana is threatened by a bacterial disease known as *Xanthomonas* wilt, which causes more than \$500 million in crop losses annually. The disease also affects other African countries in the Great Lakes region. There are no resistant varieties of banana and no ways of treating them chemically. Ugandan scientists are working on creating a GM banana using genes from sweet pepper (*Capsicum annuum*) that helps to control *Xanthomonas*.

But even if Ugandan scientists can produce such a banana, the country's farmers are banned from growing it. Only 29 countries worldwide grow GM crops, and only three of those are African. But this is set to change in coming years, according to the non-profit organization International Service for the Acquisition of Agri-biotech Applications. Kenya has already adopted a law that permits the import of GM foods. This has opened the door for approving the cultivation of GM crops.

Critics of biotechnology are right to question its role in world agriculture — like other agricultural practices, biotechnology is not risk free. Concerns such as the transfer of GM genes to wild relatives and the development of resistance to pests need to be taken seriously and kept under constant review. But addressing them requires greater investment in science and technology, not less. Such monitoring should be part of wider strategies to conserve biological



► diversity and protect human health; they should not be designed to discriminate against GM crops.

Furthermore, GM critics are wrong to conclude that because biotechnology does not solve all problems, it has no place in helping humanity to address long-term food needs. They often base their arguments on the presumption that most of the unintended consequences of genetic modification are likely to be negative, such as cases in which herbicide-resistant crops have spread to neighbouring lands. But we must look at the relative risks of all technologies, including the risks of not adopting the technology — such as how African nations would have suffered without the Green Revolution. According to a 2010 European Commission report on GM organisms, *A Decade of EU-Funded GMO Research*, based on 130 research projects spanning more than 25 years and involving more than 500 independent research groups, “biotechnology, and in particular GMOs, are not *per se* more risky than e.g. conventional plant breeding technologies”.

The application of biotechnology has a number of unintended ecological benefits. For example, the adoption of GM crops has curbed the release of greenhouse-gas emissions by reducing the use of pesticides, which require energy to manufacture. It has also saved farmers from heavy exposure to these chemicals. In addition, the use of herbicide-tolerant crops enables farmers to cut back on the ploughing and weeding that releases carbon that would otherwise be sequestered in the soil. For 2009, it is estimated that biotech crops resulted in 17.6 billion kilograms of carbon dioxide sequestration and forgone release — the equivalent of removing 7.8 million cars from the road.

Solving world hunger will involve more than just producing more food. But excluding technological options that raise productivity will do more harm than good. The international community would be better served by taking a pragmatic approach that accommodates the best available technological options, rather than relying on ideological political positions that will put the world's most vulnerable people at risk. All technological options for meeting global food needs should therefore be on the table, including agricultural biotechnology. ■

**Calestous Juma** is the director of the *Agricultural Innovation in Africa Project* at Harvard Kennedy School, Cambridge, Massachusetts 02138, USA. He is author of *The New Harvest: Agricultural Innovation in Africa* (Oxford Univ. Press, 2011).  
e-mail: calestous\_juma@harvard.edu



Giving power back to the people is key to ensuring the security of food supplies.

## Preventing hunger: Change economic policy

Simply giving people food is not enough to prevent famine, says **Peter Rosset**. Instead, we need to overhaul the policies that have upended the food supply.

**T**he global food system is broken. The number of hungry and undernourished people in the world hovers at around 1 billion<sup>1</sup> and the past few years have seen both worldwide food riots as well as epidemics of obesity and diabetes.

Fifty years ago, the United Nations World Food Programme was formed to help reduce hunger. But its original mandate of handing out food was a band-aid at best — and can actually make people more vulnerable to hunger. We now have a food system that has been destroyed by decades of misguided policies that emphasized

exports over feeding domestic populations and by runaway financial speculation. We now need to reverse those policies and fix what's broken.

According to the economic law of comparative advantage, agribusinesses should export the food, agrofuels and other products that are grown in a country, while cheaper foods are imported to feed the people. Any gaps in such a 'productionist' and 'free trade' system should then be covered by food aid, in which organizations such as the US Agency for International Development (USAID) and the World Food Programme

F. COURBET/PANOS PICTURES





Hedy Lamarr co-invented a technology to prevent remote-controlled torpedoes from being jammed.

## TECHNOLOGY

# Inventing beauty

Robert P. Crease revels in the life of a Hollywood goddess who pioneered wireless technology.

**H**edy Lamarr was no mere Hollywood starlet. The twentieth-century Austrian-American actress was also a tech-head, taking inspiration from the self-playing 'player piano' to create technology of a type now used in mobile phones. Richard Rhodes's biography, *Hedy's Folly*, gives this side of her story its due.

For some, her intelligence sits uneasily with her exquisite looks, and her invention looks like an eccentric sideline. Another biography — *Hedy Lamarr* by cinematic scholar Ruth Barton, published last year — was a solid portrait but devoted just ten pages to Lamarr's undercover life as an inventor. However, her engineering work, conducted largely in collaboration with avant-garde US composer and pianist George Antheil, was genuine. The duo patented a wireless technology to prevent jamming of

remote-controlled torpedoes. This was the first use of a system that, thanks to the particular way in which it allowed multiple users to share a common band, would later be incorporated into Wi-Fi, Bluetooth and most cordless phones.

An award-winning author on the history of the US atomic-weapons programme, Rhodes puts Lamarr's inventive spirit into coherent context. Despite its title, the book is nearly as much about Antheil: he spent the 1920s in Europe writing controversial compositions for player pianos — an early form of entertainment centre involving a primitive version of digital control. Antheil's 'orchestras' incorporated other far-out instruments: saws, hammers, electric bells, sirens and even a pair of aeroplane propellers.

**NATURE.COM**  
For more on women  
and patents:  
[go.nature.com/t5u5kp](http://go.nature.com/t5u5kp)



**Hedy's Folly:  
The Life and  
Breakthrough  
Inventions of Hedy  
Lamarr, the Most  
Beautiful Woman  
in the World**

RICHARD RHODES  
Doubleday: 2011.  
272 pp. \$26.95

Rhodes is not as good at conjuring the cultural climate as he is at describing technologies, so the book only really takes off when Lamarr meets Antheil in Hollywood in 1940. But the celebrity-riddled parties, openings and European premieres of Lamarr's day job were hardly a drab background to her hidden hobby, and her progression to stardom makes for fascinating reading.

Born in 1913 to a Jewish family in Vienna as Hedwig Kiesler, she grew into a beautiful teenager and was soon appearing in German films. The erotically charged 1933 Czech film *Ecstasy* made her a sensation: in it, she swims backstroke nude in a shimmering lake, and appears, famously, with bare breasts.

That summer, Kiesler married the first of her six husbands, a wealthy arms merchant who entertained German and Austrian weapons developers. Neither he nor his guests seemed to appreciate that their gorgeous hostess could follow conversations about submarine torpedoes and remote-control devices. When her husband tried to make her give up acting, she divorced him.

Kiesler moved to Hollywood, became Hedy Lamarr, and was soon a ravishing starlet in films such as *Algiers* (1938), the trailer for which contained the line (uttered by French actor Charles Boyer): "Come with me to zee Casbah!" But she was proud of her mental acuity. "Any girl can be glamorous," she said. "All you have to do is stand still and look stupid."

During the Second World War, Lamarr played glamour roles in films such as *Ziegfeld Girl* (1941), in which she wore a jewelled peacock-feather headdress. Meanwhile, she and Antheil had been hard at work applying for patents on weapons-related devices.

Their meeting and collaboration came at a fortuitous time for Antheil. He had been down on his luck, reduced to writing popular articles for *Esquire* magazine advising readers to evaluate women based on their glands. Lamarr hoped to tap into his expertise to find a way to augment her breasts. Antheil couldn't help her, but the two discovered a shared passion for pianos and for inventing. (Rhodes finds it "highly unlikely" that they were lovers.)

Lamarr happened to mention an idea she'd had about how to prevent jamming of remote-controlled torpedoes. Rhodes puts this simply: "If a radio transmitter and receiver are synchronized to change their

tuning simultaneously, hopping together randomly from frequency to frequency, then the radio signal passing between them cannot be jammed.” This process is now referred to as ‘frequency hopping’ or ‘spread spectrum’. Experience with player pianos had made Antheil proficient at getting machines to communicate in synchrony, so Lamarr recruited him to provide a proof of principle.

Using available documents and interviews, Rhodes retraces the steps that guided Lamarr in her thinking, and the genesis and development of the pair’s other inventions. He explains why the US Navy had no immediate use for the patent: its existing torpedoes were so problematic that it had no interest in developing another system, and Navy technicians misunderstood aspects of the Lamarr–Antheil design. So it was filed away. Rhodes charts the later, independent developments of the technology and its subsequent uses, notably in the Sonobuoy (a sonar system in a buoy) developed by the US Navy. The patent expired the year Antheil died, in 1959, when it was still classified.

The technology the two invented emerged from secrecy in 1976 and found wide application in commercial communications. In 1997, the inventors were given (Antheil posthumously) a Pioneer Award by the Electronic Frontier Foundation, a non-profit communications industry group based in San Francisco, California. But from the end of the 1940s, Lamarr’s Hollywood career had begun to drift down an all-too-familiar path. Her film performances had grown hammier, as in the overwrought and pretentious *Samson and Delilah* (1949). She gradually acquired a reputation reminiscent of *Sunset Boulevard*’s Norma Desmond, was arrested twice for shoplifting and died in 2000.

Although the book is mainly about her collaboration with Antheil, its cover sports a collage of Lamarr the starlet sitting on a long golden torpedo. Yet again we are asked to marvel at the spectacle of actress–inventor as oddity. Does this reflect sexism, the disbelief that beauty is compatible with intelligence? Would we be less surprised if her on-screen persona had been less erotic and more like Katharine Hepburn’s? Or perhaps there is a deeper cultural prejudice at work — the platonic notion that the skill involved in imitating the world is incompatible with that of understanding it.

Inside its cover, Rhodes’s book gives us the whole Hedy — a closet geek in peacock feathers — and makes that mix believable. Now it is up to us to figure out why we find that hard to digest. ■

**Robert P. Crease** is a philosopher at Stony Brook University, New York, USA. He is a columnist for *Physics World*, and his latest book is *World in the Balance*.  
e-mail: rcrease@notes.cc.sunysb.edu

## Books in brief



### Mushroom

*Nicholas P. Money* OXFORD UNIVERSITY PRESS 224 pp. \$24.95 (2011)  
Botanist Nicholas Money is unashamedly in thrall to the ‘fungal sex organ’. In this brilliant scientific and cultural exploration, these organisms of rot and soil positively sparkle. From biology to medicine, cuisine and recreation, this is a history as convoluted as the systems of fungal filaments that enrich woodlands. A human and mycological cast of thousands throngs the pages — from a 10-kilometre-square colony of honey fungus in Oregon’s Malheur National Forest, to Charles McIlvaine, author of *One Thousand American Fungi*, who fearlessly chomped his way through many of them.



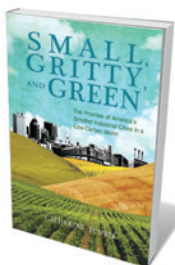
### Science on Ice: Four Polar Expeditions

*Chris Linder* UNIVERSITY OF CHICAGO PRESS 288 pp. £26 (2011)  
A century of polar science has seen vast change, not least in the researchers still braving the blizzards and gelid waters. To give an idea of their daily realities, oceanographer and photographer Chris Linder and several embedded journalists followed four Arctic and Antarctic expeditions studying, variously, an Adelie penguin colony, the Bering Sea in spring, the Greenland ice sheet and ocean pack ice in the eastern Arctic. The vivid images — of fishing for zooplankton at dawn, intent ice-breaking crews, Ernest Shackleton’s hut at Cape Royds, for instance — enliven a detailed yet accessible chronicle.



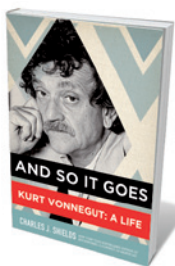
### Memory: Fragments of a Modern History

*Alison Winter* UNIVERSITY OF CHICAGO PRESS 312 pp. £19.50 (2011)  
Notions of how memory works have shifted wildly over time. Historian Alison Winter traces the evolution of memory sciences through ‘fragments’, or flashbulb moments. Drawing on sources from neurological research to diaries, she shows how the understanding of memory has deepened, ramified and sometimes taken wrong turns. Case studies include the extreme brain surgery performed by Canadian neurosurgeon Wilder Penfield; forensic hypnosis (as used in the book *The Manchurian Candidate*); psychologist Frederic Bartlett’s studies of remembering; and false-memory syndrome.



### Small, Gritty, and Green: The Promise of America’s Smaller Industrial Cities in a Low-Carbon World

*Catherine Tumber* MIT PRESS 192 pp. \$24.95 (2011)  
Detroit in Michigan may be Motown — an intrepid city founded on car manufacture — but depopulation, ill-conceived infrastructure and the flight of industry have left it devastated. Yet historian and journalist Catherine Tumber sees such urban wastelands as tomorrow’s sustainability hubs. Low population density, proximity to farmland, and a skilled workforce could aid the advent of renewable-energy technology. Plucking ideas from 25 small cities in the US Rust Belt, Tumber outlines a plausible route to a ‘repurposed’ future.



### And So It Goes: Kurt Vonnegut, A Life

*Charles J. Shields* HENRY HOLT 544 pp. £20 (2011)  
The late Kurt Vonnegut carved out his own literary landscape — blending dystopian tendencies, pitch-dark humour, autobiography and elements of his grounding in chemistry. Authorized biographer Charles J. Shields’s exhaustive research does justice to him. Whether it is the dimension-hopping Trafalmadorians, ice-nine (a solid water) or the shenanigans of fictional sci-fi writer Kilgore Trout, Vonnegut used the fantastical to comment astutely on the human condition. His honesty, Shields reminds us, is still needed.





*Micro* is the last of the posthumous publications from Michael Crichton, pictured here in 1993.

## FICTION

# Small wonder

Paul McEuen savours a technothriller from the late Michael Crichton that makes the tiny terrifying.

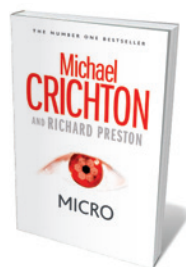
For most people, small doesn't inspire fear. We run from tanks, not integrated circuits. For a novelist writing a technothriller called *Micro*, this could be a problem. But not, it seems, for the godfather of the genre, Michael Crichton.

Crichton died in 2008, and this — following *Pirate Latitudes* (HarperCollins, 2009) — is the last of his novels to be published posthumously. Discovered as an unfinished manuscript with copious notes, it was fleshed out by best-selling author Richard Preston, whose *The Hot Zone* (Anchor, 1994) is one of the best non-fiction thrillers ever written. The result of this joint work, *Micro*, is a fast-paced mix of adventure and cutting-edge science, a novel that takes us deep into a fascinating, if violent, microscale universe.

The set-up is classic Crichton: seven graduate science students are flown to a dangerous spot on the Hawaiian island of Oahu, where a cutting-edge technology company is up to no good. Nanigen is a start-up armed with a breakthrough way to harvest pharmaceutical compounds from the Hawaiian rainforests.

Each student brings his or her own expertise: Peter Jansen is a specialist in venomous creatures; Jenny Linn a pheromone researcher; Karen King a spider expert skilled in martial arts. Also along for the ride is Danny Minot, a postmodernist all-scientific-truth-is-relative sort who is doomed for the roughest treatment. (Remember the preening, pro-environment do-gooder in Crichton's 2004 thriller *State of Fear*? He was eaten by cannibals. Danny fares much worse.)

The book's promoters tell us that the students are "transformed" and set loose to survive in the Hawaiian rainforest. Spoiler alert: I'm going to tell you what "transformed" means. Stop reading



**Micro**  
MICHAEL CRICHTON  
AND RICHARD  
PRESTON  
HarperCollins: 2011.  
448 pp. \$28.99,  
£18.99

if you don't want to know.

Crichton miniaturizes them. *Micro* is the 1989 film *Honey, I Shrunk the Kids* re-imagined as a dark, edge-of-the-seat technological thriller. The shrinking is accomplished by the powers of magnetic tensor fields, whatever those are. Even the characters, post-miniaturization, can't make sense of it: did their atoms shrink, too? How did it work?

But let that go. Give Crichton and Preston their leap. Now you are experiencing the micro-universe first-hand. The result is scary and fun — a wild romp through the Hawaiian hinterland with both man and insect on your tail.

The authors give us a potpourri of facts about chemical ecology, as well as useful hints for surviving when small (katydid 'meat' tastes like sushi). To their credit, Crichton and Preston try to keep the science as close to truth as fiction will allow.

Our heroes find everything from biomechanics to wind resistance dramatically modified, giving them, for example, ant-like super-strength, and rendering them invulnerable to long falls. This gives our protagonists a chance against predators such as insects and birds, with their terrifying arsenals of physical and chemical weapons. It is quite an adventure, pitting the mini-grads against vicious ants and a mynah bird. There is even a life-or-death fight with a bat.

That said, the book lacks some of the elements that made Crichton's best works soar. Where the technology of *Jurassic Park* (Knopf, 1990) stretched, but did not break, scientific credulity, *Micro*'s tensor-field shrinking machine is so 'out there' that it dampens the feeling that the plot could really happen. Furthermore, the story is not driven by a larger peril. If our band of students escape the microworld, they live. If not, they die. There is no ticking time bomb in the real world that makes us root for them even if we don't particularly like them.

Still, it is wonderful to see Crichton's genius — the audacious mix of adventure and techno-speak that made him famous — on vivid display one last time.

If Crichton's work had a recurring theme, it was that complex systems go awry, and that nature cannot be controlled. When Crichton was diagnosed with throat cancer in early 2008, he was expected to recover fully. The cancer had other plans. To borrow a line from *Micro*, Crichton climbed suddenly and unexpectedly into the voracious night. ■

**Paul McEuen** is professor of physics at Cornell University, Ithaca, New York 14853, USA, and director of the Kavli Institute at Cornell for Nanoscale Science. His debut thriller is *Spiral*.  
e-mail: paul@paulmceuen.com

R. O'CONNELL/F. PARESCE/E. YOUNG/WFC3 SCIENCE OVERSIGHT  
CTTE/HUBBLE HERITAGE TEAM (STSC/AURA/ESA/NASA)

Images from the Hubble Space Telescope inspire poet Tracy K. Smith, whose father helped to build it.

## Q&A Tracy K. Smith

# The space poet

Tracy K. Smith has her head in the stars. Thanks to her late father's job as an engineer on the Hubble Space Telescope, the US poet gathers inspiration from astrophysics and cosmology. Published this year, her third collection, *Life on Mars*, explores the future of human life, the great beyond and her father's death. As she prepares for a poetry reading at the Space Telescope Science Institute in Baltimore, Maryland, Smith talks about the limits of space and time.

### How did you begin to write about space?

I wrote a poem called 'Sci-Fi' several years ago that offered a clean and glamorous vision of the distant future. I felt fearless, writing "the word sun will have been re-assigned / to a Standard Uranium-Neutralizing device / found in households and nursing homes". Then my father was diagnosed with a terminal illness. My sense of the future became very personal. *Life on Mars* became a way to move towards my father, to try to understand some part of the mystery of death.



### How did your father influence you?

I always thought of him as a Renaissance man. He drew, wrote poetry, was curious about the way the world worked and how things grew. He took me to estuaries to look at bird life. He would eat a piece of fruit, dry out the seeds and plant them in the back yard. When we watched wildlife documentaries and I'd get upset about animals killing one another, he'd remind me that predation was part of a natural cycle. Now that my father is gone, I can't look up and out without feeling like I'm looking through his eyes.

### Life on Mars

TRACY K. SMITH  
*Graywolf*: 2011.  
75 pp. \$15

Several poems refer to Hubble images. Why? They come from the edge of the observable Universe. They make me feel that I'm seeing the future, even though it is actually the distant past. Either way, that impulse to career towards what is just beyond view feels similar to the way that curiosity and desire work in our everyday lives. It is like driving at night. We have this insatiable need to get beyond the edge of what we can see. There is always something out of view that the imagination races to fill in. I also have a private attachment to the Hubble because it was part of my father's daily life for a number of years. I think of it as an extension of him when I write in 'My God, It's Full of Stars':

The first few pictures came back blurred, and I felt ashamed  
For all the cheerful engineers, my father and his tribe.  
The second time,  
The optics jibed. We saw to the edge of all there is—  
So brutal and alive it seemed to comprehend us  
back.

In one poem you write that "perhaps the great error is to believe we're alone". What are your thoughts on alien life?

A friend once argued that there are probably lots of extraterrestrial life forms but at

such a distance from us that, by the time we managed to reach them, they would no longer exist. It seemed beautifully tragic to me that we may have countless neighbours calling out into the darkness, but that time is keeping us from being a part of that conversation.

### Are your poems informed by science fiction?

In Stanley Kubrick's majestic film *2001: A Space Odyssey*, his most suspenseful moments are the slowest and quietest, and his associative leaps have been instructive. I have been influenced by the visual sensibilities of classic 1970s sci-fi films, such as *The Andromeda Strain* and *The Omega Man*. Isaac Asimov's 1956 short story *The Last Question* also proved to be an elegant model for the way that imagination might defy certain hard-and-fast divisions. In it, a supreme, immaterial artificial intelligence expands in scope until every human mind is an extension of it. This intelligence comes to exist outside space and time, which fall away with the demise of the Universe. It exists solely to ponder the last question that remains: 'Is there a way to begin again?'

### Why do some of your poems imagine a dark future for human civilization?

The poems are playing with the psychology of progress. Our imperfections help to make us empathetic. My poem 'Sci-Fi' presents a sanitized vision of the future, an unsettling utopia in which we have outlived every threat, but where much of what we thrive on as humans has been deemed unnecessary. 'The Museum of Obsolescence' is about a museum far in the future where everything we no longer need as a species is housed: money, oil, books, tools. But I also imagined that some of the simplest and seemingly inconsequential things might be eternal, like vendors hawking bargain T-shirts outside.

### An earlier poem was based on the Flores human fossil. How did that inspire you?

In 2004 I read an article in *Nature* about a species of human that inhabited the Indonesian island of Flores until about 18,000 years ago. Endemic island dwarfism meant that they and everything in their environment existed on a miniature scale — there were even dwarf elephants. I wondered what it would have been like to be a woman in that place and time. I wrote the poem in the voice of someone whose emotions, desires and will to survive are the same size as ours.

### What poems are you working on next?

I have been thinking about climate change and the factions of people who doubt that it exists. I'm interested in exploring how the state of the planet may also shed light on the ways in which we treat one another. ■

INTERVIEW BY JASCHA HOFFMAN

T. RUSINGER/ROLEX/RSA



# Correspondence

## Mekong dam: invest in public health

Construction of the Xayaburi dam on the Mekong River in Laos could have serious implications for public health as well as for fishing (*Nature* **478**, 305–307; 2011).

The river and its basin are the natural habitat of the freshwater snail *Neotricula aperta*. This is the only known intermediate host of the trematode parasite *Schistosoma mekongi*, a flatworm that causes human schistosomiasis (bilharzia) through contact with infected water.

Newly created upstream water reservoirs, as well as altered river currents and sedimentation, markedly affect the relative distribution of snails carrying different *Schistosoma* species. When the Aswan High Dam was built in Egypt in 1970, it caused a shift in the relative frequency of schistosomiasis from the urinary (*S. haematobium*) to the hepatosplenic type (*S. mansoni*), which is more severe (M. F. Abdel-Wahab *et al.* *Lancet* **314**, 242–244; 1979).

Construction of the Three Gorges Dam across the Yangtze River in China is also expected to extend the snail's normal habitat substantially, increasing the likelihood of disease in domestic livestock and humans (D. P. McManus *et al.* *Clin. Microbiol. Rev.* **23**, 442–466; 2010).

The only cost-effective drug currently available to treat infection is the schistosomicide praziquantel. However, parasite resistance to this drug is emerging as a result of mass administration programmes and high rates of reinfection in endemic areas (S. D. Melman *et al.* *PLoS Negl. Trop. Dis.* **3**, e504; 2009).

A plea on behalf of poor people living along the Mekong's banks is unlikely to affect the construction of economically important projects such as the Xayaburi dam. Instead, some

of the dam's revenues should be invested in the development of new antischistosomal drugs and vaccines.

**Antoni R. Blaazer** *Bilharzia Foundation, Wormerveer, the Netherlands. ton.blaazer@bilharziafoundation.org*

## Mekong dam: assess health risks

Huge hydroelectric projects (see, for example, *Nature* **478**, 305–307; 2011) can have a negative impact on public health, particularly in tropical areas. Public and private power companies should be obliged to assess this risk and incorporate it into their development policies.

It was pointed out ten years ago that the World Commission on Dams, created in 1998 by the World Bank and the International Union for Conservation of Nature, had fallen short in analysing health issues (A. C. Sleight and S. Jackson *Lancet* **357**, 570–571; 2001). This oversight has still not been rectified: the Strategic Environmental Assessment of Hydropower on the Mekong Mainstream, for example, did not include public-health risks or the likely consequences in its report.

The failure to develop strategies to mitigate emerging or re-emerging infectious diseases in this context probably stems from ignorance and financial considerations. Reliable analysis of environmental and societal health risks is essential for their prevention and control, which depend on evidence-based guidance for health policy and planning. It is also crucial to optimize integrated surveillance systems to allow timely responses to public-health threats.

The 1,070-megawatt Nam Theun 2 hydropower station in central Laos has shown that long-term surveillance and monitoring of vector-borne diseases is achievable at an acceptable cost. **Gilles Guerrier** *Paris, France. guerriergilles@gmail.com*

## Space telescope: debt problems go deeper

You report concern among some US planetary scientists that the funding for the James Webb Space Telescope (JWST) is preventing NASA from providing a rocket for the 2016 ExoMars launch (see [go.nature.com/zimumq](http://go.nature.com/zimumq) and [go.nature.com/emqasy](http://go.nature.com/emqasy)). But this is not the root of the problem.

The real issue is the dismal five-year budget projection for NASA's Solar System exploration in US President Barack Obama's budget proposal from February 2011 — a projection driven by debt-reduction pressures across all federal agencies.

A look at more recent versions of the 2012 budget — one from the House of Representatives, which provides no money for the JWST, and the other from the Senate, which gives enough this year to support a launch in 2018 — reveals that identical amounts are provided for Solar System exploration. It is therefore clear that cancellation of the JWST will lead to debt reduction and not to a transfer of funds to planetary missions, a point made explicitly in the House budget text.

The JWST will peer back in time to the beginnings of the cosmos while measuring the composition of atmospheres of super-Earths around nearby stars. The challenge for space scientists is to make the case that such exploration is worthwhile, even in difficult economic times. If we attack one another's programmes, we shall see funds for many of our hard-fought missions swept away in the name of deficit reduction — and we shall all lose out.

**Jonathan I. Lunine** *Cornell University, Ithaca, New York, USA. jlunine@astro.cornell.edu*

## Space telescope: focus on priorities

The concerns and frustration of the planetary community over the uncertainty of the 2016 ExoMars mission are

understandable (see [go.nature.com/zimumq](http://go.nature.com/zimumq)). Several budgetary factors are to blame, and these precede and are unrelated to the recently increased costs of launching the James Webb Space Telescope.

The US President's 2012 budget proposal for NASA's Planetary Science Division, released in February 2011, dropped from the previous year's figure by roughly US\$1 billion over the next five years.

In addition, the cost of the Mars Science Laboratory, initially estimated at \$650 million in 2003 (at a time when all major mission costs were underestimates), has risen to around \$2.5 billion.

The cost of the missions in NASA's Science Mission Directorate budget for 2008 was also found to be understated by billions of dollars.

The scientific community needs to support the decadal priorities of each of the science divisions at NASA and avoid firing inwards, which will threaten funding for science overall.

**Garth D. Illingworth** *University of California Santa Cruz, California, USA. gdi@ucolick.org*

## Sea-ice loss sparks oil drilling boom

An irony that should give us all reason to pause is that the unprecedented, climate-induced loss of Arctic sea ice (*Nature* **478**, 171; 2011) has triggered an explosion of oil exploration off Greenland and in the Kara, Barents and Chukchi seas.

**James B. McClintock** *University of Alabama at Birmingham, Alabama, USA. mcclinto@uab.edu*

### CONTRIBUTIONS

See author guidelines at <http://go.nature.com/cmchno>.

# Does gravity correct gauge couplings?

ARISING FROM D. J. Toms *Nature* **468**, 56–59 (2010)

Whether various models of quantum gravity predict observable effects is a matter of dispute. Toms has argued<sup>1</sup> that there are quantum gravity corrections to the energy dependence of the electric charge in quantum electrodynamics (QED) that depend quadratically on the energy, and cause the electric charge to vanish at high energies. This conclusion was based on a background field calculation using a cut-off thought to be related to energy. We argue that scattering processes cannot have such a quadratic energy dependence. Hence the quadratic correction<sup>1</sup> cannot apply to a charge that is physically measurable, and does not lead to asymptotic freedom in QED.

In the absence of gravitation, couplings in gauge field theories (such as the electric charge in QED) vary logarithmically with energy due to quantum corrections, with that in QED increasing. It has been suggested<sup>2</sup> that quantum gravity would make a correction that depends quadratically on energy, causing all gauge couplings to vanish at high energies (even that in QED), the property known as asymptotic freedom. An observation of this effect would have profound implications for the unification of gravity with the gauge interactions, as it would cause the latter's couplings to drop very rapidly at scales above the expected grand unification scale,  $10^{16}$  GeV.

However, several confusing and contradictory calculations of such gravitational corrections have appeared subsequently, ranging from explicit gauge dependence<sup>3</sup> to the absence of such terms in dimensional regularization<sup>4</sup>, contradicting results in other approaches<sup>5</sup>.

A covariant approach to the problem was taken recently by Toms<sup>1</sup>, based on a gauge-invariant heat-kernel regularization using a generalized background-field method. It was claimed that the quadratic energy dependence of the gravitational contributions to the electric charge was confirmed, and an additional logarithmic energy dependence was found, proportional to a positive cosmological constant. This calculation used a proper time cut-off  $\tau_c$  related to an energy cut-off  $E_c$ , which was then identified with the energy  $E$  at which the renormalized electric charge  $e(E)$  is evaluated. The resulting corrections would render asymptotically free the corresponding QED running coupling, and similar results would hold for non-Abelian gauge theories.

The disagreements<sup>1–5</sup> raise the question of whether the claimed gravitational corrections to gauge couplings are physical, and specifically the question of whether the appearance of a quadratic cut-off dependence actually signals the appearance of an asymptotically-free coupling in physical processes: the previous calculation<sup>1</sup> could be absorbed into a trivial charge renormalization if  $E_c$  and  $E$  were not identified.

Physical measurements—for example, of the electric charge—are derived from on-shell scattering amplitudes (S-matrix elements). These are invariant under local redefinitions of fields, as shown in the equivalence theorem<sup>6,7</sup>. As we now show, the equivalence theorem implies that energy-dependent modifications of gauge couplings, such as those discussed above, cannot affect S-matrix elements, and hence are not relevant for asymptotic freedom or the unification of gauge interactions with gravity—for example, in string theory.

The equivalence theorem<sup>6,7</sup> asserts that if one redefines a generic field  $\varphi$  by  $\varphi \rightarrow \Phi = \varphi + F(\varphi)$ , where  $F(\varphi)$  is a local, gauge invariant combination of  $\varphi$  and its derivatives that does not influence the mass-shell condition for the particle associated with  $\varphi$ , and the correlation functions of  $F$  with itself and/or  $\varphi$  do not have poles corresponding to massless particles, then the S-matrix in the transformed theory is the same as in the original theory.

As the claimed gravitational corrections to gauge couplings may be removed by a field redefinition satisfying the conditions of the

equivalence theorem, these corrections can have no physical effects on on-shell scattering processes. This also explains the discrepancies described above, including the apparent dependences on the gauge-fixing parameter and the regularization scheme, because the quantities being calculated were not physically measurable.

From the point of view of an effective action, the  $E^2$ -dependent terms<sup>1–5</sup> correspond to higher-derivative terms of the form:  $\frac{b}{M_P^2} \int d^4x F_{\mu\nu} \square F^{\mu\nu}$ , where  $b$  is a dimensionless numerical constant,  $M_P \approx 10^{19}$  GeV is the Planck energy scale,  $F_{\mu\nu}$  is the electromagnetic field strength and  $\square$  is a covariant second derivative. Terms of the form  $(\nabla_\rho F_{\nu\mu})^2$  can easily be cast in the same form, which is the only independent higher-derivative combination that is quadratic in the field strength and in space-time derivatives, and terms of the form  $\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu}$  vanish thanks to the cyclic permutation identity. It is straightforward to see that the coefficient  $b$  above can be changed by the following local field redefinition of the gauge potential,  $A_\mu$ , which respects the criteria of the equivalence theorem outlined above:  $A_\mu \rightarrow \tilde{A}_\mu = A_\mu + \frac{c}{M_P^2} \nabla^\nu F_{\nu\mu}$ , where  $\tilde{A}_\mu$  denotes the redefinition of  $A_\mu$ ,  $c$  is an arbitrary numerical constant and  $\nabla_\mu$  denotes a gravitationally-covariant derivative. All corrections to the photon propagator can be removed by such local redefinitions, and no terms with any number of derivatives that are bilinear in the gauge potential  $A_\mu$ , such as  $(\nabla_\rho \dots \nabla_\lambda F_{\alpha\beta})^2$  and so on, have any effect on on-shell scattering amplitudes, as is well known in string theory<sup>8</sup>.

It follows that there are no relevant power-law gravitational corrections to the physical electric charge, no asymptotic freedom in QED, and no effect on the comparison between gauge and gravitational interaction strengths in string unification scenarios, as had been shown explicitly in the context of open strings<sup>9,10</sup>, without appeal to the equivalence theorem. This observation can be extended to terms of higher orders in the effective gauge theory, but does not extend to the term proportional to the cosmological constant that depends logarithmically on the energy scale, which cannot be absorbed by a local field redefinition.

A related point of view was made recently in the context of a scalar field theory with a  $\lambda\Phi^4$  interaction, where it was argued that power-law corrections due to quantum gravitational interactions do not signify a running of the coupling constant  $\lambda$  that can be measured physically<sup>11</sup>. In the context of our discussion, this is another illustration of the equivalence theorem.

We conclude by restating our principal conclusion: the equivalence theorem implies that S-matrix elements are unaffected by higher-order derivative corrections that are quadratic in the gauge fields, and hence that the measurable electric charge does not exhibit a quadratic energy dependence leading to asymptotic freedom, as suggested by Toms<sup>1</sup>.

John Ellis<sup>1</sup> & Nick E. Mavromatos<sup>2</sup>

<sup>1</sup>Theory Division, CERN, CH-1211 Geneva 23, Switzerland.

email: john.ellis@cern.ch

<sup>2</sup>Department of Physics, King's College London, Strand, London WC2R 2LS, UK.

Received 10 January; accepted 2 September 2011.

1. Toms, D. J. Quantum gravitational contributions to quantum electrodynamics. *Nature* **468**, 56–59 (2010).
2. Robinson, S. P. & Wilczek, F. Gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* **96**, 231601 (2006).



3. Pietrykowski, A. R. Gauge dependence of gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* **98**, 061801 (2007).
4. Toms, D. J. Quantum gravity and charge renormalization. *Phys. Rev. D* **76**, 045015 (2007).
5. Tang, Y. & Wu, Y. L. Quantum gravitational contributions to gauge field theories. *Commun. Theor. Phys.* **54**, 1040–1044 (2010).
6. Salam, A. & Strathdee, J. A. Equivalent formulations of massive vector field theories. *Phys. Rev. D* **2**, 2869–2876 (1970).
7. Kallosh, R. E. & Tyutin, I. V. The equivalence theorem and gauge invariance in renormalizable theories. *Yad. Fiz.* **17**, 190–209 (1973).
8. Tseytlin, A. A. Ambiguity in the effective action in string theories. *Phys. Lett. B* **176**, 92–98 (1986).
9. Antoniadis, I. *et al.* Aspects of type I – type II – heterotic triality in four-dimensions. *Nucl. Phys. B* **489**, 160–178 (1997).
10. Kiritsis, E. & Kounnas, C. Infrared regularization of superstring theory and the one loop calculation of coupling constants. *Nucl. Phys. B* **442**, 472–493 (1995).
11. Anber, M. M., Donoghue, J. F. & El-Houssieny, M. Running couplings and operator mixing in the gravitational corrections to coupling constants. Preprint at (<http://arXiv.org/abs/1011.3229>) (2010).

**Author Contributions** The two authors contributed equally to this paper.

**Competing Interests** Declared none.

doi:10.1038/nature10619

## Toms replies

REPLYING TO J. Ellis & N. E. Mavromatos *Nature* **479**, doi:10.1038/nature10619 (2011)

In a previous paper<sup>1</sup>, I used the gauge-invariant background-field method to calculate a contribution to the renormalization of electric charge due to quantum gravity that was quadratically dependent on an energy cut-off. This was done by taking a constant electromagnetic background field,  $F$ . The result was used to support the original suggestion of Robinson and Wilczek<sup>2</sup> that quadratic divergences could lead to asymptotic freedom. This result has been criticized by Ellis and Mavromatos<sup>3</sup> and their present submission<sup>4</sup> gives a shortened and altered discussion of their viewpoint.

The implication of Ellis and Mavromatos<sup>4</sup> is that the quadratic divergence that I found<sup>1</sup> is related to the dimension-six operator  $F\partial^2 F$ . This is not the case. The quadratic divergence proportional to  $F^2$  that was calculated has nothing to do with  $F\partial^2 F$ . I agree with these authors that the coefficient of  $F\partial^2 F$  can be altered by a field redefinition and cannot affect any physically measurable quantity; by taking  $F$  to be constant as I did, the dimension-six operator cannot occur. What Ellis and Mavromatos<sup>4</sup> argue is that the dimension-six operator cannot affect the renormalization (and hence running) of charge, and nothing more; their criticism has no direct bearing on the calculations reported by Robinson and Wilczek<sup>2</sup> or myself<sup>1</sup>.

The reason why I do not now believe that the quadratic divergences contribute to a running electric charge has to do with not properly identifying a physically measurable definition of the charge. (I am grateful to J. Donoghue (personal communication) for pointing this out.) This should indeed follow from an S-matrix calculation but no one, including Ellis and Mavromatos, has attempted this calculation for Einstein–Maxwell (or Yang–Mills) theory. (A related S-matrix calculation has been done<sup>5</sup> for a non-gauge field.) Instead the background-field method in one form or another has been used. A

physical definition of charge can be considered completely within the background-field method where it can be shown<sup>6</sup> that the quadratic divergences do not contribute to the running electric charge, and only logarithmic divergences do so contribute. Although it does not appear that the quadratic divergences lead to a running of gauge coupling constants, it is still possible<sup>7</sup> that quantum gravity can lead to asymptotic freedom if there is a positive cosmological constant; however, the running is only logarithmic, not quadratic, and is not as interesting phenomenologically.

### D. J. Toms<sup>1</sup>

<sup>1</sup>Newcastle University, School of Mathematics and Statistics, Herschel Building, Newcastle upon Tyne NE1 7RU, UK.  
email: d.j.toms@newcastle.ac.uk

1. Toms, D. J. Quantum gravitational contributions to quantum electrodynamics. *Nature* **468**, 56–59 (2010).
2. Robinson, S. P. & Wilczek, F. Gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* **96**, 231601 (2006).
3. Ellis, J. & Mavromatos, N. E. On the interpretation of gravitational corrections to gauge couplings. Preprint at (<http://arXiv.org/abs/1012.4353>) (2010).
4. Ellis, J. & Mavromatos, N. E. Does gravity correct gauge couplings? *Nature* **479**, doi:10.1038/nature10619 (this issue).
5. Anber, M. M., Donoghue, J. F. & El-Houssieny, M. Running couplings, and operator mixing in the gravitational corrections to coupling constants. *Phys. Rev. D* **83**, 124003 (2011).
6. Toms, D. J. Quadratic divergences and quantum gravitational contributions to gauge coupling constants. *Phys. Rev. D* **84**, 084016 (2011).
7. Toms, D. J. Cosmological constant and quantum gravitational corrections to the running fine structure constant. *Phys. Rev. Lett.* **101**, 131301 (2008).

doi:10.1038/nature10620

## EARTH SCIENCE

# Probing the core's light elements

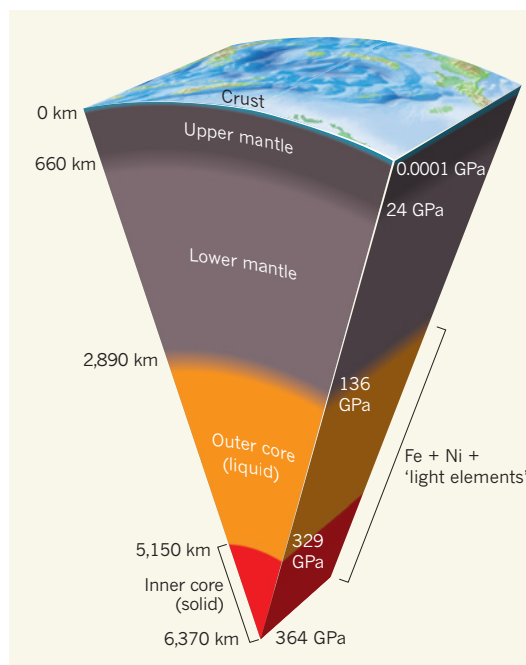
A fine marriage between seismic data and laboratory experiments carried out at the extreme conditions of Earth's deep interior indicates that the planet's liquid outer core is poor in oxygen. [SEE LETTER P.513](#)

THOMAS S. DUFFY

Earth's core accounts for one-third of the planet's mass and has a central role in Earth's overall energy budget and dynamics. Although the core has long been known to be composed mainly of iron, together with some nickel<sup>1,2</sup>, the identity of the lighter elements that make up about 8% of the core's mass has been an enigma for nearly 60 years. Work reported in this issue by Huang *et al.*<sup>3</sup> (page 513) combines the results of laboratory experiments with geophysical data for the core to address this long-standing mystery. Better knowledge of the core's main light element (or elements) will shed light on heat flow in Earth's deep interior, on the origin and growth of the core's solid inner region, and on the generation and evolution of Earth's magnetic field<sup>4</sup>.

But which element is it? Here we can bring out the line-up of usual suspects: sulphur, oxygen, silicon, carbon and hydrogen. Each is geologically abundant and can dissolve in liquid iron under the appropriate range of pressure and temperature conditions. Geochemical arguments for and against each can be constructed<sup>5</sup>. Each also has its own implications for core formation and evolution of the early Earth<sup>6</sup>. For example, an oxygen-dominated core would imply that oxidizing conditions were present during much of core formation, whereas a silicon-dominated one would require mostly reducing conditions. Hence, unmasking the identity and abundance of the core's main light elements will also be a major step forward in understanding Earth's geochemical evolution.

Seismic data<sup>7</sup> provide robust constraints on density and sound velocities throughout the core. These data reveal that the core is divided into a large, liquid outer core with a radius of about 3,500 kilometres, within which is embedded a smaller, solid inner core with a radius of about 1,200 km (Fig. 1). The differences in seismic velocity and density between the liquid outer and solid inner cores are also known, as is the existence of weak seismic anisotropy in the inner core. One interesting



**Figure 1 | Cross-section of Earth's interior.** Depths for the various regions are indicated on the left and corresponding pressures (in gigapascals) on the right. The core is mainly composed of iron (Fe) and nickel (Ni), with some unknown lighter elements. By comparing laboratory measurements with seismic data, Huang *et al.*<sup>3</sup> place new constraints on the identity of these light elements in the core.

feature is that the abundance of light elements in the inner core seems to be only about half of that in the outer<sup>8</sup>. All of these observable aspects provide important characteristics that the successful light-element candidates must have. However, directly probing potential core materials against these geophysical criteria is challenging because of the extreme conditions that exist in the core — pressures of 136–364 gigapascals and temperatures of about 4,000–6,500 kelvin.

Shock-compression experiments, which subject samples to high-velocity impactors, provide one of the best routes to simulating core conditions in the laboratory. These are dynamic experiments that generate very high pressures for microsecond durations. One advantage of shock-wave experiments is that high temperatures are necessarily generated along with high pressure. For samples with

iron-rich compositions, the pressure–temperature states achieved under shock loading are fortuitously close to those expected to occur in the core<sup>9,10</sup>. Although the temperatures are somewhat uncertain, the need for large extrapolations of laboratory results to core conditions, as is common for room-temperature experiments using diamond anvil cells for example, is avoided. The shock experiments are also conducted under conditions in which the sample has been transformed into the liquid state and thus are more directly comparable to the liquid outer core.

Shock-wave experiments traditionally measure the density achieved under high-pressure loading. However, density data alone cannot provide a unique constraint on core composition. Huang *et al.*<sup>3</sup> overcome this problem by also measuring the sound velocity of the shocked material. By comparing measurements of both density and sound velocity with the seismic data for the core, it finally becomes possible to affirm or rule out certain compositions.

Huang and colleagues used one oxygen-rich and sulphur-poor (8 weight per cent O and 2 wt% S) composition for their experiments and one oxygen-poor and sulphur-rich composition (2.2 wt% O and 5.3 wt% S). They found that the oxygen-rich composition yields sound velocities substantially higher than those of the liquid outer core. For compositions within the iron–sulphur–oxygen system, those with more than about 2.5 wt% oxygen are not able to simultaneously match both the density and sound-velocity profiles of the core. By contrast, they showed that sulphur-rich compositions can indeed match both density and sound velocity. Does this mean that sulphur is the dominant light element? Not necessarily. Models of Earth's accretion have generally concluded that the amount of sulphur in the core must be small<sup>11</sup>. Therefore, it is essential to perform similar experiments on a wider range of components. As noted previously, carbon and silicon are both plausible candidates on geochemical grounds, and so experiments on these materials are needed.

Further efforts to nail down core composition would require examining how potential



light elements partition between the core's solid and liquid portions at its temperature and pressure conditions. This is necessary to explain the compositional differences between the solid and liquid parts of the core. A theoretical study<sup>12</sup> addressing this question suggests that oxygen — not sulphur — may have the right partitioning behaviour, in conflict with Huang and colleagues' experimental results<sup>3</sup>. The extreme conditions of the core make a laboratory test of this partitioning behaviour highly challenging, but experimental capabilities are making rapid advances on several fronts in high-pressure-temperature science. The past few decades have seen remarkable

progress in uncovering the state and structure of many features of the deep Earth, with the light elements of the core remaining a stubborn puzzle. The work reported here suggests that a solution to the problem may finally be at hand. ■

**Thomas S. Duffy** is in the Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA.  
e-mail: duffy@princeton.edu

1. Birch, F. J. *Geophys. Res.* **57**, 227–286 (1952).
2. McQueen, R. G. & Marsh, S. P. *J. Geophys. Res.* **71**, 1751–1756 (1966).
3. Huang, H. *et al. Nature* **479**, 513–516 (2011).

4. Buffett, B. A. *et al. J. Geophys. Res.* **101**, 7989–8006 (1996).
5. Poirier, J.-P. *Phys. Earth Planet. Inter.* **85**, 319–337 (1994).
6. Wood, B. J., Walter, M. J. & Wade, J. *Nature* **441**, 825–833 (2006).
7. Kennett, B. L. N., Engdahl, E. R. & Buland, R. *Geophys. J. Int.* **122**, 108–124 (1995).
8. Jephcoat, A. & Olson, P. *Nature* **325**, 332–335 (1987).
9. Brown, J. M. & McQueen, R. G. *J. Geophys. Res.* **91**, 7485–7494 (1986).
10. Huang, H. *et al. J. Geophys. Res.* **115**, B05207 (2010).
11. Dreibus, G. & Palme, H. *Geochim. Cosmochim. Acta* **60**, 1125–1130 (1996).
12. Alfè, D., Gillan, M. J. & Price, G. D. *Earth Planet. Sci. Lett.* **195**, 91–98 (2002).

## CANCER

# Final act of senescence

**Damaged cells can initiate cancer. To avert this, faulty cells disable their own propagation by undergoing senescence. But for full protection against liver cancer, the senescent cells must be cleared by the immune system. SEE LETTER P.547**

MANUEL SERRANO

Cellular damage is unavoidable and intrinsic to life, and it is often aggravated by pathogens or exposure to toxic agents. In long-lived multicellular organisms, however, damaged cells pose a problem not because of their loss of function, but mainly because their proliferation can escalate a minor local problem into a deadly cancer. Fortunately, cells can detect their own level of damage and eventually dismiss themselves in a final act of altruism for the benefit of the organism. Self-dismissal comes in two forms, apoptosis (self-destruction) and senescence (self-disabling), with the latter being at least as prominent as the former in providing protection against cancer<sup>1,2</sup>. On page 547 of this issue, Kang *et al.*<sup>3</sup> explore how the immune system targets senescent cells to limit the development of liver cancer.

The fate of senescent cells within tissues has long been uncertain. Some show extremely lengthy residency times *in vivo*. For instance, moles formed by senescent melanocyte cells can reside in human skin for decades<sup>4</sup>. Until recently, therefore, the possibility that senescent cells could be selectively targeted for elimination was generally overlooked. This changed with the discovery that such cells secrete signalling molecules that contribute to the establishment of senescence<sup>5,6</sup> and promote tissue repair by inducing proliferation of neighbouring cells and attracting cells of the immune system<sup>7</sup>.

Rapid clearance of senescent cells was first observed in mouse liver cancers. In this work<sup>8</sup>, reactivation of a silent *p53* gene (a

major inducer of senescence) made cancer cells 'aware' of their own damage, triggering a strong senescence response. The cancer then became a senescent mass and was swiftly and selectively eliminated by immune cells — a fate delivered mainly by the innate arm of the immune system (macrophages, neutrophils and natural killer cells) (Fig. 1).

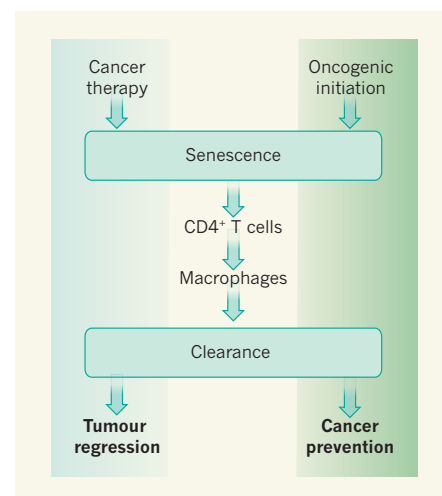
Subsequent work<sup>9</sup> reached similar conclusions. Disabling the driving oncogenes in mouse lymphoma and leukaemia resulted in tumour regression that was associated with apoptosis and senescence. In these cancer models, the immune system was necessary for the induction of senescence — but not apoptosis — and for complete and long-term tumour remission. The essential immune cells were CD4<sup>+</sup> T cells<sup>9</sup>, which are part of the adaptive immune system and can recognize cancer cells. In the absence of CD4<sup>+</sup> T cells, senescence was not induced and tumour remission was incomplete, being rapidly followed by tumour relapse.

These studies<sup>8,9</sup> offered insight into how cancers respond to therapies. But understanding the role of the immune system in preventing the emergence of cancer is equally important. Kang *et al.*<sup>3</sup> set out to do just that. The authors introduced the *Nras*<sup>G12V</sup> oncogene into some 15% of liver cells in mice. Six days later, these 'transduced' — but otherwise normal — cells naturally became senescent and were progressively cleared during the following weeks, becoming undetectable after 2 months.

In a veritable tour de force, Kang and colleagues demonstrate that clearance depended on CD4<sup>+</sup> T cells — which showed specificity towards the *Nras*<sup>G12V</sup>

mutation — and that it required senescence. Mouse liver cells that had been transduced with a non-oncogenic *Nras* gene (*Nras*<sup>G12V/D38A</sup>) did not undergo senescence or clearance 2 months later. Moreover, transduction of *Nras*<sup>G12V</sup> into mice lacking the crucial senescence mediator Arf did not trigger a specific immune response against the *Nras*<sup>G12V</sup> mutation. These observations imply that the T-cell-mediated immune response to oncogenic mutations requires senescence. Although CD4<sup>+</sup> T cells were necessary for clearance of senescent cells, they instigated recruitment of macrophages, which directly executed the clearance.

What is most exciting about this work is



**Figure 1 | Clearance of senescent cells.** When damaged, cells may undergo senescence to disable their own proliferation and may even trigger their own clearance. In the case of cancer, treatments that lead to senescence and clearance by immune cells — CD4<sup>+</sup> T cells and macrophages — cause tumour regression<sup>8,9</sup>. Kang *et al.*<sup>3</sup> find that mouse liver cells that carry the hyperactive *Nras*<sup>G12V</sup> oncogene, but are otherwise normal, also trigger senescence and clearance by the immune system, and that these processes are essential for complete protection from cancer. CD4<sup>+</sup> T cells are essential for instigating clearance in some<sup>3,9</sup>, but not all<sup>8</sup>, cases, whereas macrophages are the final common executors of clearance.

that the clearance of senescent cells leads to protection against cancer. In fact, when *Nras*<sup>G12V</sup>-expressing senescent cells were not cleared (for example, in mice lacking CD4<sup>+</sup> T cells), numerous tumours appeared, whereas clearance of senescent cells robustly prevented tumour development.

The authors<sup>3</sup> also studied mouse models of liver inflammation and toxicity — known risk factors for liver cancer in humans. Again, in the absence of the adaptive immune system, cells that had undergone senescence following inflammation and liver toxicity were not cleared, and this was associated with increased tumorigenesis.

Kang *et al.* further extrapolated their findings to human disease, by analysing the livers of patients infected with hepatitis C virus (HCV) whose immune systems were either intact or compromised owing to concomitant infection with HIV. In agreement with their mouse data, the authors found that senescent cells were cleared in patients infected with HCV alone but not in those infected with

both HCV and HIV. Similarly, HCV-infected patients with cirrhosis who were being treated with immunosuppressants had senescent cells in their livers, but this was not the case in untreated patients. As the authors note, HIV infection and immunosuppressant therapy are associated with increased risk of liver cancer.

It has been shown<sup>10</sup> that clearance of senescent cells by the immune system is also important for the resolution of fibrotic scars in liver. Together with Kang and co-workers' data, this suggests that the clearance of these cells could indeed prove generally beneficial for tissue repair, as previously proposed<sup>7</sup>.

This paper<sup>3</sup> clearly shows that cell clearance is essential for the maximum tumour-suppressive effects of senescence. But this brings us to a conundrum: why do senescent cells, which cannot proliferate, pose a cancer risk? Two possible answers come to mind. First, it is conceivable that rare senescent cells exist at an intermediate — still undefined — state that has the potential to resume proliferation and initiate cancer. Second, the

proliferative factors secreted by senescent cells for their tissue-repair function may be usurped by neighbouring cancerous cells to enhance their proliferation. Future work will undoubtedly resolve this issue and will continue to enrich the fascinating link between senescence and cancer. ■

**Manuel Serrano** is at the Spanish National Cancer Research Centre (CNIO), Madrid E28029, Spain.

e-mail: mserrano@cnio.es

1. Collado, M. & Serrano, M. *Nature Rev. Cancer* **10**, 51–57 (2010).
2. Hanahan, D. & Weinberg, R. A. *Cell* **144**, 646–674 (2011).
3. Kang, T.-W. *et al.* *Nature* **479**, 547–551 (2011).
4. Michaloglou, C. *et al.* *Nature* **436**, 720–724 (2005).
5. Acosta, J. C. & Gil, J. *Cancer Res.* **69**, 2167–2170 (2009).
6. Kuilman, T. & Peeper, D. S. *Nature Rev. Cancer* **9**, 81–94 (2009).
7. Campisi, J. *Curr. Opin. Genet. Dev.* **21**, 107–112 (2011).
8. Xue, W. *et al.* *Nature* **445**, 656–660 (2007).
9. Rakhra, K. *et al.* *Cancer Cell* **18**, 485–498 (2010).
10. Krizhanovsky, V. *et al.* *Cell* **134**, 657–667 (2008).

## INTERFACIAL CHEMISTRY

# Gold's enigmatic surface

**Gold is not as inert as was believed — it can promote molecular synthesis. A study uses scanning tunnelling microscopy to catch gold in the act as it guides the formation of one-dimensional polymers from saturated hydrocarbons.**

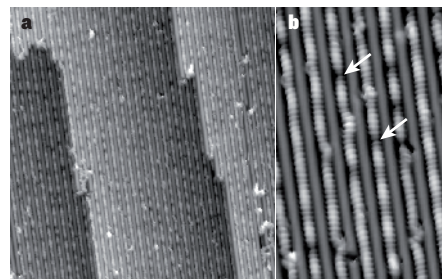
ROBERT J. MADIX & CYNTHIA M. FRIEND

Carbon-carbon bond-forming reactions are used to make molecules that have specific structures and properties, such as polymers. Particularly challenging is the use of saturated hydrocarbons (alkanes) as starting materials for these reactions, because strong carbon-hydrogen (C–H) bonds in the molecules must first be broken. Harsh reaction conditions are often required to do this, but in general such conditions make it difficult to break a specific C–H bond and thereby selectively control where in the molecule carbon-carbon bond formation occurs. Writing in *Science*, Zhong *et al.*<sup>1</sup> report that the adsorption of alkanes on a gold surface changes the structure of one-dimensional channels in the metal. These channels direct the polymerization of the alkanes, so that carbon-carbon bonds form only at, or close to, the ends of the molecules.

This result might sound surprising, because gold has traditionally been believed to be chemically inert. However, research has shown that the metal's surface is not inert, and it is now

used as a supporting surface for self-assembled monolayers of molecules<sup>2</sup>. Furthermore, special reaction centres can be created on gold, for example by binding oxygen atoms to it, and this has facilitated the development of gold-catalysed reactions for making carbon-oxygen and carbon-nitrogen bonds<sup>3–5</sup>. Such reactions have been used to selectively catalyse the formation of low-molecular-weight compounds that are important in the chemical industry<sup>6</sup>. The advantage of using gold to promote reactions is that, where the reaction centres are not present, it is less able to break chemical bonds than are other commonly used catalytic metals. This means that, if a primary reaction can be initiated on the surface, the intermediate chemical species created will be unlikely to take part in side reactions, thus providing excellent control over which products form.

Zhong and colleagues' reaction<sup>1</sup> illustrates the use of such effects to make compounds of high molecular weight. The authors predicated their experiments on two principles. The first was that high-molecular-weight alkanes remain bound to gold surfaces at temperatures that are high enough to break C–H bonds. The



**Figure 1 | Polymerization on gold.** Zhong *et al.*<sup>1</sup> report that alkane monomers on a gold surface react selectively at their chain ends to form polymers. **a**, In this micrograph, the polymer chains (bright lines) line up in parallel along troughs in the gold surface. The area shown is 50 nm × 50 nm. **b**, This inset shows the chains at higher magnification. The arrows indicate gaps in the rows of gold atoms that underlie the polymers. These gaps form during the polymerization reaction, but the role of the missing atoms in the reaction, if any, is unknown. The area shown is 9 nm × 13 nm.

second principle was that, once C–H bonds of an alkane have ruptured to form alkyl intermediates on a gold surface, recombination of neighbouring alkyl groups will occur, yielding polymer chains.

The rupture of C–H bonds by metallic gold is well documented: for instance, butyl groups (CH<sub>3</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>) attached to gold surfaces undergo C–H bond cleavage at 280 kelvin to form butene<sup>7</sup> (CH<sub>3</sub>CH<sub>2</sub>CH=CH<sub>2</sub>); and terminal C–H bonds in *t*-butyl thiol bound to gold break at 490 K (ref. 8; *t*-butyl thiol is (CH<sub>3</sub>)<sub>3</sub>CSH). Furthermore, alkanes bind to the surfaces of precious metals primarily by van der Waals interactions, which means that the strength of binding increases with the number of carbon atoms in the alkane chain (and so with



molecular weight). Although the temperatures that are needed to desorb alkanes from gold have not been measured, it is known<sup>9</sup> that a temperature of 550 K is required to desorb the long alkane C<sub>32</sub>H<sub>66</sub> from graphite. Because the forces that bind them to gold, this is indicative of the temperatures needed to desorb C<sub>32</sub>H<sub>66</sub> from the metal. It is therefore clear that long alkanes should remain attached to a gold surface at sufficiently high temperatures for C–H bond cleavage to occur, as Zhong *et al.* assumed.

There is also evidence to support the second principle. When heated, neighbouring alkyl groups (derived from compounds known as alkyl bromides) couple together on gold surfaces to form alkanes<sup>7</sup>. Furthermore, scanning tunnelling microscopy has previously been used to show that carbon–carbon bonds form between large, complex organic molecular fragments on gold to yield highly ordered surface structures<sup>10</sup>. So, overall, the two principles are well enough understood to be predictive of the reactions reported by the authors<sup>1</sup>.

But perhaps the most intriguing aspect of Zhong and colleagues' work is the alignment of long alkanes along the troughs of the gold surface, and the associated reorganization of the surface structure that widens the troughs. This alignment may be key to the observed selective cleavage of the terminal C–H bonds and might also facilitate coupling of the resulting alkyl groups. There is precedent for such alignment: acetylene (HC≡CH) reacts with surface-bound oxygen atoms on silver surfaces to yield adsorbed monomers, which polymerize in such a way that their carbon–carbon triple bonds are confined within troughs<sup>11</sup>. Silver and gold are in the same group of the periodic table, and they have the same surface structure, so Zhong and colleagues' findings<sup>1</sup> add to the growing evidence of the commonality in the chemical behaviour of these surfaces.

But there are also striking differences in the behaviour of silver and gold. In the previous work on silver<sup>11</sup>, each monomer was connected by an intervening silver atom, so that the resulting polymer was formed from a repeated –(C–C–Ag)– atom motif. Zhong *et al.* obtained images of their polymers on gold surfaces using scanning tunnelling microscopy (Fig. 1), and saw no indication that gold atoms were incorporated. The authors did, however, observe vacancies in the gold surfaces, where gold atoms had disappeared during the reaction. A lingering question is the fate of these vanished atoms and their role in the polymerization process. On the face of it, the results suggest that interactions between polymers and rows of gold atoms on the surface of the metal provide sufficient force to reconstruct the surface.

Further experimental and theoretical studies should be performed to address two crucial points about the polymerization — the role of alkane confinement in the channels and the role of surface reconstruction. Perhaps

confinement of molecules in the channels makes gold atoms along ridges in the metal surface accessible to the terminal C–H bonds in the alkane chains. Experiments in which the molecular structure and chain length of the alkanes are varied, along with molecular-dynamics simulations, would cast light on this issue, and provide deeper insight into these fascinating coupling reactions. ■

**Robert J. Madix** is at the School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. **Cynthia M. Friend** is at the SLAC National Accelerator Laboratory, Stanford University, Stanford, California 94305, USA. e-mails: rmadix@seas.harvard.edu; cfriend@slac.stanford.edu

#### PALAEOANTHROPOLOGY

## The earliest modern humans in Europe

The reanalysis of findings from two archaeological sites calls for a reassessment of when modern humans settled in Europe, and of Neanderthal cultural achievements. **SEE LETTERS P.521 & P.525**

PAUL MELLARS

On page 521 of this issue, Higham *et al.*<sup>1</sup> document the presence of modern humans at the extreme northwestern tip of Europe by around 42,000–43,000 years ago — based on the most recent calibration of the radiocarbon timescale<sup>2</sup>. And on page 525, Benazzi *et al.*<sup>3</sup> report a previously unsuspected, very early dispersal of modern humans into the extreme southern fringes of Europe by around 43,000–45,000 years ago. The two papers shed new light on the earliest settlement of two widely separated regions of Europe by anatomically and genetically modern human (*Homo sapiens*) populations, following their initial dispersal from Africa about 50,000 to 60,000 years ago.

The two papers<sup>1,3</sup> crucially depend on the same premise — that one can differentiate clearly and reliably between the remains of modern humans and those of the immediately preceding Neanderthal populations on the basis of the shape, dimensions, cusp patterns, enamel thickness and other features of human teeth<sup>4</sup>. Although attributions based on these features may be questioned by some, both papers present clear and well-documented justifications for their taxonomic identifications, and will probably be accepted by the majority of specialists in the field.

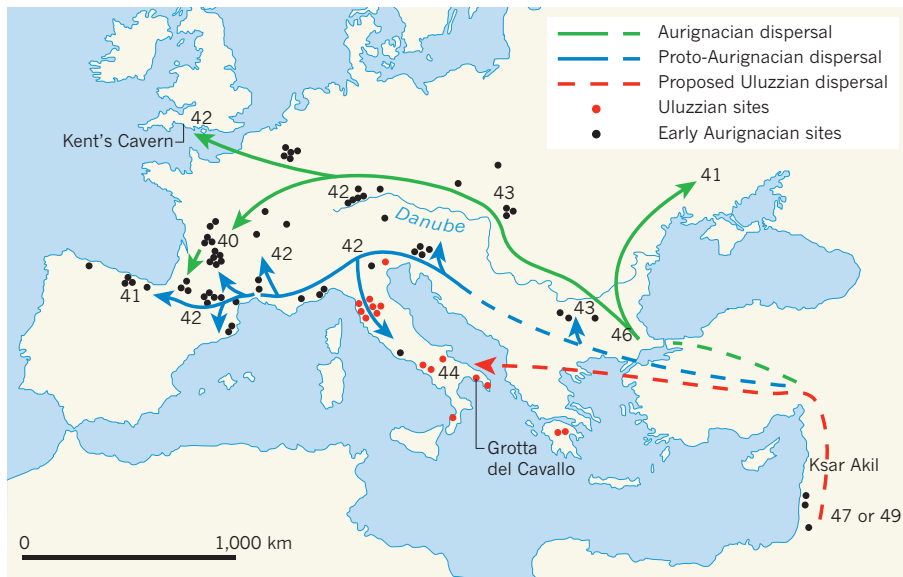
Higham and colleagues' evidence<sup>1</sup> comes

1. Zhong, D. *et al.* *Science* **334**, 213–216 (2011).
2. Bain, C. D. *et al.* *J. Am. Chem. Soc.* **111**, 321–335 (1989).
3. Xu, B., Zhou, L., Madix, R. J. & Friend, C. M. *Angew. Chem.* **122**, 404–408 (2010).
4. Wittstock, A., Zielasek, V., Biener, J., Friend, C. M. & Bäumer, M. *Science* **327**, 319–322 (2010).
5. Angelici, R. J. *J. Organometallic Chem.* **693**, 847–856 (2008).
6. Christensen, C. H. & Nørskov, J. K. *Science* **327**, 278–279 (2010).
7. Paul, A., Yang, M. X. & Bent, B. E. *Surf. Sci.* **297**, 327–344 (1993).
8. Jaffey, D. M. & Madix, R. J. *J. Am. Chem. Soc.* **116**, 3012–3019 (1994).
9. Paserba, K. R. & Gellman, A. J. *Phys. Rev. Lett.* **86**, 4338–4341 (2001).
10. Grill, L. *et al.* *Nature Nanotechnol.* **2**, 687–691 (2007).
11. Guo, X.-C. & Madix, R. J. *Surf. Sci.* **564**, 21–28 (2004).

from a site in the United Kingdom — Kent's Cavern in south Devon. It consists of a fragment of a human maxilla — part of the upper jaw — originally discovered during excavations on the site in 1927 (Fig. 1). The bone was provisionally attributed to *H. sapiens* by the anthropologist Arthur Keith on the basis of its overall anatomical form<sup>5</sup>. Higham *et al.* systematically compared the dimensions of the three teeth in the maxilla with well-documented specimens of Neanderthal and modern human teeth from other European sites. Their findings reinforce the attribution



**Figure 1 | Human remains.** Higham and colleagues' reconstruction<sup>1</sup> of a maxilla — part of an upper jawbone — found at Kent's Cavern, UK, confirms that the remains belong to an anatomically modern human. (Taken from Supplementary Information to ref. 1.)



**Figure 2 | Principal dispersal routes of the earliest modern humans across Europe.** Aurignacian populations of modern humans spread mainly along the Danube valley, whereas Proto-Aurignacian populations spread along the Mediterranean coast. Both apparently originated from the Near East (based, for example, on findings at Ksar Akil, Lebanon<sup>14,17</sup>). Dashed lines indicate less-certain dispersal routes, and numbers represent dates of sites in thousands of years before present (BP), based on the most recent calibration of the radiocarbon timescale<sup>2</sup>. Higham *et al.*<sup>1</sup> report that bones from a site at Kent's Cavern, UK, are dated to between 42,000 and 43,000 years BP, which makes them the earliest dated remains of anatomically modern humans in western Europe. Another early modern human culture, known as the Uluzzian industry, is currently confined to sites in Italy and Greece<sup>9</sup>. Benazzi *et al.*<sup>3</sup> date Uluzzian shell samples from a site at Grotta del Cavallo, Italy, to between 43,000 and 45,000 years BP. This suggests that a previously unknown modern human dispersal occurred across southern Europe. (Graphic, Dora Kemp.)

of the maxilla to fully modern humans.

The authors also attempted direct radiocarbon dating of one of the teeth, but this failed because the collagen in the tooth was poorly preserved. They therefore systematically dated a long series of animal bone samples recovered stratigraphically above and below the reported position of the maxilla at Kent's Cavern, using the newly developed 'ultrafiltration' method for eliminating chemical contamination from the bones. Their results suggest that the maxilla itself dates from between 42,000 and 43,000 years before present (BP), making it the earliest dated specimen of an anatomically modern human in western Europe.

The idea that anatomically modern humans had penetrated western Europe by at least 42,000 years ago is strongly borne out by radiocarbon dating of several archaeological sites that belong to the Aurignacian culture in western and central Europe<sup>6,7</sup> (Fig. 2), which is now generally accepted to have been a modern human culture<sup>4,8</sup>, rather than a Neanderthal one. By contrast, Benazzi and colleagues' radiocarbon dating<sup>3</sup> of finds from the cave site of Grotta del Cavallo in the extreme southern 'heel' of Italy demonstrates an entirely new and hitherto unsuspected component of the modern human settlement of southern Europe.

The authors analysed a series of eight marine shell samples recovered in close stratigraphic association with two human teeth from the middle levels of the site. The remains

were associated with a highly distinctive form of stone-tool culture known as the Uluzzian industry, which is geographically confined to sites in the Italian peninsula, apart from two recently discovered sites in southern Greece<sup>9</sup> (Fig. 2). Benazzi and colleagues' radiocarbon dating of the shell samples — which works by exclusively analysing the aragonite components of the shells, rather than the calcite — points to an age for these samples in the range of 43,000–45,000 years old. They could even be slightly older, if the possibility of some residual chemical contamination of the samples is taken into account.

Uluzzian technology combines an apparent mixture of both Middle Palaeolithic and Upper Palaeolithic elements, including several simple, shaped bone tools and a number of perforated or tubular seashell ornaments<sup>9</sup>. Up to now, this has been widely accepted as the result of acculturation — the exchange of cultural features — between indigenous Neanderthal populations in Italy and intrusive, early 'Proto-Aurignacian' populations of modern humans on the adjacent Mediterranean coast<sup>10–12</sup> (Fig. 2). This in turn implies that the humans responsible for the Uluzzian culture were Neanderthals, rather than modern humans. But on the basis of their morphological analysis<sup>3</sup> of the two teeth associated with the Grotta del Cavallo shell samples, Benazzi *et al.* conclude that the teeth are distinctively modern human, and not Neanderthal. If so, this casts radically new light

on the true significance of the much-debated Uluzzian industries of the Italian peninsula. As the authors point out, this evidence<sup>3</sup> further weakens previously proposed arguments<sup>10,11</sup> that European Neanderthals independently invented many distinctively 'modern' cultural features without any contacts with or input from intrusive modern human populations<sup>12</sup>.

The most intriguing issue raised by Benazzi and colleagues' new interpretation of the Uluzzian culture focuses on the demographic origins of the very early modern human populations in this extreme southern fringe of Europe, and is likely to generate lively debate amongst archaeologists and anthropologists. If one assumes that the earliest modern human populations entered Europe through the Levantine corridor from Africa (as all the current genetic evidence suggests<sup>13</sup>), then presumably one must look to sites in the Near East for the immediate ancestors of the Uluzzian populations (Fig. 2).

The most relevant site in this context is Ksar Akil, Lebanon, where occasional 'crescentic' or 'lunate' forms of stone tools characteristic of the Italian Uluzzian industries have been found in archaeological levels associated with some of the earliest modern human occupations<sup>14</sup>. Similar forms are also recorded in several rather earlier (about 50,000–60,000 years BP) sites in East Africa<sup>15</sup>, from which the initial dispersal of modern humans from Africa to Eurasia is generally agreed to have derived. Intriguingly, an analogous, pre-Aurignacian dispersal of modern humans into parts of Central Europe from a Near Eastern source may also have occurred: tool assemblages belonging to the Bohunician and Bacho-Kirian cultures<sup>9,16</sup> have been discovered at sites such as Bacho-Kiro and Temnata in Bulgaria, and at Bohunice and Stránská Skála in the eastern Czech Republic, with radiocarbon dates of broadly similar age to those of the Uluzzian levels at the Grotta del Cavallo.

In any event, the unexpectedly early dating of anatomically modern human remains in Italy<sup>3</sup> poses challenging questions for future studies of the earliest dispersal of modern humans into the more southerly, Mediterranean fringes of Europe. Exactly where did these populations originate? What demographic or climatic factors propelled their expansion from east to west? And how did they successfully displace the immediately preceding Neanderthal populations in the Italian peninsula and elsewhere? But for now, the two papers<sup>1,3</sup> published in this issue effectively demonstrate how new techniques and analyses can radically change previously held views on the character and timing of the earliest modern human colonization of Europe. ■

**Paul Mellars** is in the Department of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK.  
e-mail: pam59@cam.ac.uk



1. Higham, T. *et al.* *Nature* **479**, 521–524 (2011).
2. Reimer, P. J. *et al.* *Radiocarbon* **51**, 1111–1150 (2009).
3. Benazzi, S. *et al.* *Nature* **479**, 525–528 (2011).
4. Bailey, S. E., Weaver, T. D. & Hublin, J.-J. *J. Hum. Evol.* **57**, 11–26 (2009).
5. Keith, A. *Trans. Proc. Torquay Nat. Hist. Soc.* **5**, 1–2 (1927).
6. Schmidt, C. C., Brou, L. & Jaccottet, L. *J. Archaeol. Sci.* **37**, 3320–3337 (2010).
7. Nigst, P. R. in *When Neanderthals and Modern Humans Met* (ed. Conard, N. J.) 269–304 (Kerns, 2006).
8. Trinkaus, E. *Annu. Rev. Anthropol.* **34**, 207–230 (2005).
9. Kozłowski, J. K. in *Rethinking the Human Revolution* (eds Mellars, P., Boyle, K., Stringer, C. & Bar-Yosef, O.) 317–328 (McDonald Inst., 2007).
10. D'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pelegrin, J. *Curr. Anthropol.* **39**, S1–S44 (1998).
11. D'Errico, F. *Evol. Anthropol.* **12**, 188–202 (2003).
12. Mellars, P. *Evol. Anthropol.* **14**, 12–27 (2005).
13. Soares, P. *et al.* *Curr. Biol.* **20**, R174–R183 (2010).
14. Azoury, I. *Ksar Akil, Lebanon* (Br. Archaeol. Rep., 1986).
15. Ambrose, S. H. *J. Archaeol. Sci.* **25**, 377–392 (1998).
16. Svoboda, J. A. in *Rethinking the Human Revolution* (eds Mellars, P., Boyle, K., Stringer, C. & Bar-Yosef, O.) 329–339 (McDonald Inst., 2007).
17. Mellars, P. *Evol. Anthropol.* **15**, 167–182 (2006).

## PLANETARY SCIENCE

# Europa awakening

**Brines percolating in the icy crust of Jupiter's moon Europa may be responsible for the satellite's enigmatic chaotic terrains. A new model predicts that one such terrain is currently forming over shallow subsurface water. SEE LETTER P.502**

LASZLO P. KESZTHELYI

On page 502 of this issue, Schmidt and colleagues<sup>1</sup> posit that Europa, one of the four large satellites of Jupiter, is currently geologically active\*. This result could profoundly affect the future exploration of Europa, which is consistently NASA's top target for robotic exploration in the outer Solar System<sup>2,3</sup>. With an ocean of liquid water roughly twice the volume of Earth's oceans lying beneath an icy crust, Europa is the most obvious place to search for extraterrestrial life<sup>2,3</sup>. Although the ultimate goal of exploring Europa is to sample the ocean, the next space mission is focused on gauging the thickness of the frozen carapace that another mission will need to pierce. The overwhelming majority of researchers agree that the ice crust is approximately  $20 \pm 10$  kilometres thick<sup>4–6</sup>; however, the appearance of an unresolved controversy is maintained by a handful of studies that suggest the crust is only a few kilometres thick<sup>7,8</sup>. This debate has remained frozen by the absence of new observational constraints since the demise of the Galileo orbiter in 2003.

Schmidt and colleagues<sup>1</sup> shake this impasse by coupling insights from terrestrial glacial processes with refined analyses of the Galileo data. The authors did not discover a place on Earth that mimics the alien conditions of Europa. Instead, they built on a new recognition of the role of brines (salt water) along the margins of the Antarctic ice sheet and improved understanding of the movement of water and ice in response to subglacial volcanic eruptions. Specifically, salt water can flow through fractures in ice, transporting significant volumes of both water and dissolved minerals over large distances. When

these brines do eventually freeze, material has been redistributed and the stresses within the ice have been rearranged. Also, an investigation<sup>9</sup> of subglacial eruptions in locations such as Iceland showed that the stresses around a small body of liquid water within a large body of ice serve to keep the liquid confined in a discrete, subsurface, lens-shaped region.

Combining these ideas, Schmidt *et al.*<sup>1</sup> propose a model for the formation of 'chaos terrain' on Europa. A chaos is a discrete patch of the surface consisting of disrupted crustal blocks that have a superficial resemblance to icebergs calving into the sea. The model relies on a plume of warm, convecting, pure ice bringing heat up to a layer of salty ice that has a lower melting temperature. When this salty ice melts, there is a reduction in its volume and the surface collapses. The resulting fractures are filled with brine that eventually freezes and expands, uplifting the matrix between large collapsed crustal blocks. This model fits the available observations better than previous models. Most interestingly, it suggests that the chaos at a surface feature on Europa known as Thera Macula, which is about 100 kilometres in diameter, is currently forming over liquid brine existing within a few kilometres of the surface.

The authors' suggestion<sup>1</sup> that Europa has internal processes that are active today is important for many different reasons. Although previous studies have shown that the surface of Europa is geologically young, with an average age of around 30 million to

70 million years old<sup>10</sup>, current geological activity has not been detected<sup>11</sup>. If active, Europa would be in the exclusive company of Earth, Jupiter's moon Io and Saturn's moon Enceladus. The activity observed on Mars, Saturn's moon Titan and Neptune's moon Triton seems to be driven by solar energy rather than internal heat.

Furthermore, Schmidt and colleagues' study provides an important target for future missions to Europa. If their model is correct, Thera Macula should have changed markedly in the decades between Galileo and the next mission to Europa. Such changes would mark a location that has a pocket of water at a depth that is much more accessible than the underlying main ocean.

As a final note, this study could well be a blueprint for how breakthroughs in planetary science will continue even in a period of fiscal austerity and fewer new data. There is still much to be done to extract information from older data and then convert that into scientific understanding. For example, the addition of topographic information to two-dimensional surface images and the spectroscopic identification of minerals have often proved invaluable in choosing between different models for subsurface processes. However, because the derivation of topography and mineralogy from the available data is still complex and time-consuming, a vast treasure trove has yet to be mined.

Similarly, the days of making discoveries by simply visually comparing features on Earth with extraterrestrial ones are probably drawing to a close. Instead, as in this study, it will be essential to understand fundamental processes on Earth and then extrapolate to how these processes interact in an alien setting. Thus, in many ways, Schmidt and colleagues' work could well be illuminating the path for future exploration of not just Europa but the wider Solar System. ■

**Laszlo P. Keszthelyi** is at the *Astrogeology Science Center, US Geological Survey, Flagstaff, Arizona 86001, USA.*  
e-mail: laz@usgs.gov

1. Schmidt, B. E., Blankenship, D. D., Patterson, G. W. & Schenk, P. M. *Nature* **479**, 502–505 (2011).
2. National Research Council *New Frontiers in the Solar System: An Integrated Exploration Strategy* (National Academies Press, 2003).
3. National Research Council *Vision and Voyages for Planetary Science in the Decade 2013–2022* (National Academies Press, 2011).
4. Pappalardo, R. T. *et al.* *Nature* **391**, 365–368 (1998).
5. Turtle, E. P. & Pierazzo, E. *Science* **294**, 1326–1328 (2001).
6. McKinnon, W. B. *Geophys. Res. Lett.* **26**, 951–954 (1999).
7. Greenberg, R. G. *et al.* *Icarus* **141**, 263–286 (1999).
8. Hoppa, G. V., Tufts, B. R., Greenberg, R. & Geissler, P. E. *Science* **285**, 1899–1902 (1999).
9. Björnsson, H. *Glob. Planet. Change* **35**, 255–271 (2002).
10. Zahnle, K., Schenk, P., Levison, H. & Dones, L. *Icarus* **163**, 263–289 (2003).
11. Phillips, C. B. *et al.* *J. Geophys. Res.* **105**, 22579–22598 (2000).

\*This article and the paper<sup>1</sup> under discussion were published online on 16 November 2011.

# The genome of *Tetranychus urticae* reveals herbivorous pest adaptations

Miodrag Grbić<sup>1,2\*</sup>, Thomas Van Leeuwen<sup>3\*</sup>, Richard M. Clark<sup>4\*</sup>, Stephane Rombauts<sup>5,6</sup>, Pierre Rouzé<sup>5,6</sup>, Vojislava Grbić<sup>1,2</sup>, Edward J. Osborne<sup>4</sup>, Wannes Dermauw<sup>3</sup>, Phuong Cao Thi Ngoc<sup>5,6</sup>, Félix Ortego<sup>7</sup>, Pedro Hernández-Crespo<sup>7</sup>, Isabel Diaz<sup>8</sup>, Manuel Martinez<sup>8</sup>, Maria Navajas<sup>9</sup>, Élio Sucena<sup>10,11</sup>, Sara Magalhães<sup>12</sup>, Lisa Nagy<sup>13</sup>, Ryan M. Pace<sup>13</sup>, Sergej Djuranović<sup>14</sup>, Guy Smagghe<sup>3</sup>, Masatoshi Iga<sup>3</sup>, Olivier Christiaens<sup>3</sup>, Jan A. Veenstra<sup>15</sup>, John Ewer<sup>16</sup>, Rodrigo Mancilla Villalobos<sup>16</sup>, Jeffrey L. Hutter<sup>17</sup>, Stephen D. Hudson<sup>17</sup>, Marisela Velez<sup>18</sup>, Soojin V. Yi<sup>19</sup>, Jia Zeng<sup>19</sup>, Andre Pires-daSilva<sup>20</sup>, Fernando Roch<sup>21</sup>, Marc Cazaux<sup>1</sup>, Marie Navarro<sup>1</sup>, Vladimir Zhurov<sup>1</sup>, Gustavo Acevedo<sup>1</sup>, Anica Bjelica<sup>1</sup>, Jeffrey A. Fawcett<sup>5,6†</sup>, Eric Bonnet<sup>5,6†</sup>, Cindy Martens<sup>5,6</sup>, Guy Baele<sup>5,6</sup>, Lothar Wissler<sup>22</sup>, Aminael Sanchez-Rodriguez<sup>23</sup>, Luc Tirry<sup>3</sup>, Catherine Blais<sup>24</sup>, Kristof Demeestere<sup>25</sup>, Stefan R. Henz<sup>26</sup>, T. Ryan Gregory<sup>27</sup>, Johannes Mathieu<sup>28</sup>, Lou Verdon<sup>29</sup>, Laurent Farinelli<sup>30</sup>, Jeremy Schmutz<sup>31,32</sup>, Erika Lindquist<sup>32</sup>, René Feyereisen<sup>33</sup> & Yves Van de Peer<sup>5,6</sup>

The spider mite *Tetranychus urticae* is a cosmopolitan agricultural pest with an extensive host plant range and an extreme record of pesticide resistance. Here we present the completely sequenced and annotated spider mite genome, representing the first complete chelicerate genome. At 90 megabases *T. urticae* has the smallest sequenced arthropod genome. Compared with other arthropods, the spider mite genome shows unique changes in the hormonal environment and organization of the Hox complex, and also reveals evolutionary innovation of silk production. We find strong signatures of polyphagy and detoxification in gene families associated with feeding on different hosts and in new gene families acquired by lateral gene transfer. Deep transcriptome analysis of mites feeding on different plants shows how this pest responds to a changing host environment. The *T. urticae* genome thus offers new insights into arthropod evolution and plant–herbivore interactions, and provides unique opportunities for developing novel plant protection strategies.

Mites belong to the Chelicerata, the second largest group of terrestrial animals. Chelicerates represent a basal branch of arthropods. Subsequent to their origin in the Cambrian period, arthropods radiated into two lineages: the Chelicerata and the Mandibulata (comprising the Myriapoda and the Pancrustacea (which includes both crustaceans and insects))<sup>1,2</sup>. Extant lineages of chelicerates include Pycnogonida, Xiphosura (horseshoe crabs) and Arachnida (a large group comprising scorpions, spiders and the Acari (ticks and mites))<sup>3,4</sup> (Supplementary Fig. 1.1). Within the Acari, *T. urticae* belongs to the Acariformes with the earliest fossils dating from the Lower Devonian period (410 million years ago). The Acari represent the most diverse chelicerate clade, with over 40,000 described species that exhibit tremendous variations in lifestyle, ranging from parasitic to predatory to plant-feeding. Some mites are of major concern to

human health and include allergy-causing dust mites, scabies mites and mite vectors of scrub typhus<sup>5</sup>.

The two-spotted spider mite, *Tetranychus urticae*, is a cosmopolitan agricultural pest<sup>6</sup> belonging to an assemblage of web-spinning mites. The name ‘spider’ highlights their ability to produce silk-like webbing used to establish a colonial micro-habitat, protect against abiotic agents, shelter from predators, communicate via pheromones and provide a vehicle for dispersal<sup>7</sup>.

*Tetranychus urticae* represents one of the most polyphagous arthropod herbivores, feeding on more than 1,100 plant species belonging to more than 140 different plant families including species known to produce toxic compounds. It is a major pest in greenhouse production and field crops, destroying annual and perennial crops such as tomatoes, peppers, cucumbers, strawberries, maize, soy,

<sup>1</sup>Department of Biology, The University of Western Ontario, London N6A 5B7, Canada. <sup>2</sup>Instituto de Ciencias de la Vid y el Vino (CSIC, UR, Gobierno de La Rioja), 26006 Logroño, Spain. <sup>3</sup>Department of Crop Protection, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium. <sup>4</sup>Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA. <sup>5</sup>Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium. <sup>6</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. <sup>7</sup>Department of Environmental Biology, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain. <sup>8</sup>Centro de Biotecnología y Genómica de Plantas, UPM-INIA, 28223 Madrid, Spain. <sup>9</sup>INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus international de Baillarguet, 34988 Montpellier-sur-Lez, France. <sup>10</sup>Instituto Gulbenkian de Ciência, 2781-901 Oeiras, Portugal. <sup>11</sup>Universidade de Lisboa, Faculdade de Ciências, Departamento de Biologia Animal, 1749-016 Lisbon, Portugal. <sup>12</sup>Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental, 1749-016 Lisbon, Portugal. <sup>13</sup>Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA. <sup>14</sup>Johns Hopkins University School of Medicine, Department of Molecular Biology & Genetics, Baltimore, Maryland 21205, USA. <sup>15</sup>Institut de Neurosciences Cognitives et Intégratives d'Aquitaine Université de Bordeaux 1, 33405 Talence, France. <sup>16</sup>Centro Interdisciplinario de Neurociencia de Valparaíso, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso 2360102, Chile. <sup>17</sup>Department of Physics and Astronomy, The University of Western Ontario, N6A 5B7 London, Canada. <sup>18</sup>Instituto de Catálisis y Petroleoquímica CSIC, Madrid, Spain; IMDEA Nanociencias, Facultad de Ciencias, Universidad Autónoma de Madrid, 28050 Madrid, Spain. <sup>19</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. <sup>20</sup>Department of Biology, University of Texas at Arlington, Arlington, Texas 76019, USA. <sup>21</sup>Université de Toulouse, UPS, Centre de Biologie du Développement, Université Paul Sabatier, 31062 Toulouse, France; Centre National de la Recherche Scientifique, UMR 5547, Centre de Biologie du Développement, 31062 Toulouse, France. <sup>22</sup>Westfälische Wilhelms University, Institute for Evolution and Biodiversity, Evolutionary Bioinformatics Group, Hufferstrasse 1, D-48149 Münster, Germany. <sup>23</sup>CMPG, Department of Microbial and Molecular Systems, K.U. Leuven, B-3001 Leuven, Belgium. <sup>24</sup>UPMC Univ Paris 06, UMR CNRS 7622, Equipe Biogenèse des signaux hormonaux, Case 29, 75005 Paris, France. <sup>25</sup>Research Group EnVOC, Department of Sustainable Organic Chemistry and Technology, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium. <sup>26</sup>Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany. <sup>27</sup>Department of Integrative Biology, University of Guelph, N1G 2W1 Guelph, Canada. <sup>28</sup>Boyce Thompson Institute for Plant Research, Ithaca, New York 14853, USA. <sup>29</sup>Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, N5V 4T3 London, Canada. <sup>30</sup>Fasteris SA, CH-1228 Plan-les-Ouates, Switzerland. <sup>31</sup>HudsonAlpha Institute for Biotechnology Huntsville, Alabama 35806, USA. <sup>32</sup>DOE Joint Genome Institute, Walnut Creek, California 94598, USA. <sup>33</sup>UMR 1301, INRA, CNRS and Université de Nice Sophia Antipolis, 06903 Sophia Antipolis, France. †Present addresses: Institut Curie, 26 rue d'Ulm, Paris 75248, France; INSERM, U900, Paris 75248, France; Mines ParisTech, Fontainebleau 77300, France (E.B.); Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan (J.A.F.).

\*These authors contributed equally to this work.



apples, grapes and citrus. The recent introduction of the related species *Tetranychus evansi* to Europe and Africa from South America demonstrates the invasive nature of these pests in global agriculture<sup>8</sup>. Computer modelling suggests that with intensifying global warming, the detrimental effects of spider mites in agriculture will markedly increase<sup>9</sup> due to accelerated development at high temperatures.

*Tetranychus urticae* is known for its ability to develop rapid resistance to pesticides. Among arthropods it has the highest incidence of pesticide resistance<sup>10</sup>. Chemical control often causes a broad cross-resistance within and between pesticide classes, resulting in resistance to novel pesticides within 2–4 years. Many aspects of the biology of the spider mite, including rapid development, high fecundity and haplo-diploid sex determination, seem to facilitate rapid evolution of pesticide resistance. Control of multi-resistant mites has become increasingly difficult and the genetic basis of such resistance remains poorly understood<sup>11</sup>.

As the first completed chelicerate genome, the comparison of the *T. urticae* genome with the genomes of insects and the crustacean *Daphnia pulex* expands the arthropod genetic toolkit. At the same time, the very compact *T. urticae* genome has unique attributes among arthropod genomes with remarkable instances of gene gains and losses. The completion of the *T. urticae* genome sequence opens new avenues for understanding the fundamentals of plant–herbivore interactions, developing novel pest-management strategies and producing new biomaterials on the nanometre scale.

### The small genome of *T. urticae*

The *T. urticae* genome (strain London) was sequenced (Sanger) to 8.05× coverage and assembled into 640 scaffolds covering 89.6 megabases (Mb) (Supplementary Notes 1, 2.1 and 2.2). 70,778 Sanger expressed sequence tag (EST) sequences from embryos, larvae, nymphs and adults were generated, and further complemented with RNA-seq data on matching samples (Supplementary Note 2.3). We identified 18,414 protein-coding gene models, of which 84% (15,397) are supported by EST (8,243), protein homology (11,433) and/or RNA-seq data (14,545) (Supplementary Note 2.4 and Supplementary Fig. 2.4.1). From alignments of ~43-million paired-end Illumina reads from a second *T. urticae* strain (Montpellier) to the London sequence, 542,600 single nucleotide polymorphisms and small indels were predicted (Supplementary Note 2.5). The complete genome annotation of *T. urticae* is available at the BOGAS website<sup>12</sup>. With an estimated genome size of about 90 Mb, the *T. urticae* genome is the smallest arthropod genome sequenced so far. The genomes of other chelicerates are much larger (565–7,100 Mb), with the unfinished genome of the tick *Ixodes scapularis* estimated at 2,100 Mb<sup>13</sup>. Multiple characteristics of the *T. urticae* genome correlate with its compact size: small transposable element content and microsatellite density, increased gene density and holocentric chromosomes (see Supplementary Note 3.1 for chromosomal features).

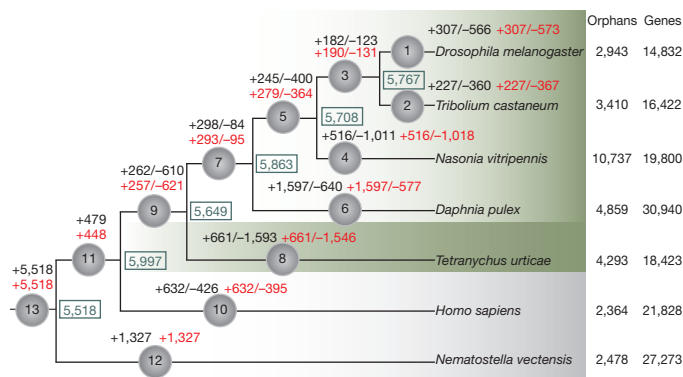
Transposable elements totalled 9.09 Mb (Supplementary Note 3.2), putting *T. urticae* together with *D. pulex* and *Apis mellifera* as arthropods with 10% or less of their genomes comprised of transposable elements. Long terminal repeat (LTR) retrotransposons, and in particular Gypsy-like elements, were the most abundant type of transposable elements. L1-like Long interspersed elements (LINEs), Tc1/Mariner-like DNA transposons, and Maverick (Polinton) elements were also detected (see Supplementary Table 3.2.1). Deep sequencing of small RNAs (~19–30 nucleotides) across developmental stages (Supplementary Note 4.1) identified 226,829 unique RNAs that mapped to 676,266 different loci in the genome. The number of unique small RNA counts per size category shows a peak at 21 and 26 nucleotides. These two peaks include short interfering RNAs and Piwi-interacting RNAs, respectively, similar to what is observed in *Drosophila melanogaster*<sup>14</sup>. Their alignments to the genome indicate that both probably silence diverse transposable elements. Included among ~21-nucleotide small RNAs are 52

predicted microRNAs (miRNAs). On the basis of the identity of their seed regions (nucleotides 2–7 of the miRNA sequence), the *T. urticae* miRNAs can be grouped into 43 families (Supplementary Note 4). Half of the predicted miRNAs were not conserved when compared to annotated miRNAs and available genomes of other arthropods<sup>15</sup>, suggesting that they might be *T. urticae*- or lineage-specific (Supplementary Tables 4.3.1–4.3.4).

The microsatellite density in the *T. urticae* genome is among the lowest observed for arthropods (Supplementary Note 3.3 and Supplementary Fig. 3.3.1), consistent with the expectation that repeat content of genomes typically scales with genome size. The *T. urticae* microsatellite classes have a distinct profile: mono-nucleotide repeats are virtually non-existent, and di-nucleotide repeats, normally the most abundant type of microsatellites, are found significantly less often than tri-nucleotides, as in *Tribolium castaneum*<sup>16</sup>. The gene density is twice as high compared to *D. melanogaster*, with 205 versus 92 genes per Mb, respectively. The mean number of exons per gene was low and similar to that found in *D. melanogaster* (~3.8 exons per gene). The size distribution of introns was typically skewed with a mean intron size of 400 bp and a median of 96 bp (see Supplementary Note 3.4, Supplementary Fig. 2.4.3 and Supplementary Table 2.4.1). The holocentric nature of *T. urticae* chromosomes<sup>17</sup> (the absence of centromeres and the diffuse nature of the kinetochores) is correlated with a lack of large tracts of gene-poor heterochromatin. The uniformly distributed gene density (Supplementary Note 3.1.1 and Supplementary Fig. 3.2.1) contrasts with the human body louse (*Pediculus humanus*, Phthiraptera, a hemimetabolous insect with a small genome), where 95% of the genes are concentrated in only 55 Mb of the 110-Mb genome<sup>18</sup>.

### Comparative genomics

As the first completely sequenced and annotated chelicerate genome, the *T. urticae* genome expands the set of arthropod genomes beyond Pancrustacea and provides an important out-group for comparative genomics. Comparison of the coding gene repertoire of *T. urticae* with the arthropods *T. castaneum*, *D. melanogaster*, *Nasonia vitripennis* and *D. pulex*, the chordate *Homo sapiens*, and the cnidarian *Nematostella vectensis* (Fig. 1) resulted in 2,667 shared gene families (Supplementary Note 5.1). Almost 3,000 gene families are common to the arthropods sampled, whereas 5,038 gene families (8,329 genes) are unique to *T. urticae* (Supplementary Fig. 5.1.1). Of those, 622 gene families (1,398 genes) have homologues in species other than those listed above, most of which belong to other arthropods. Homologues of 74 gene families (93 genes) were found in the unfinished genomes of tick<sup>13</sup> and/or *Varroa destructor*<sup>19</sup> and are probably chelicerate, rather than specific to *T. urticae*. Therefore, 4,416 gene families (6,609 genes) were found to be unique to *T. urticae*. A gene gain/loss analysis (Fig. 1 and



**Figure 1 | Gene family history.** At each time point (grey circles), the number of gains (+) and losses (-) of gene families is indicated as inferred by DOLLOP (black) and CAFE (red) programs. The inferred ancestral number of gene families, according to DOLLOP, is shown in green boxes.

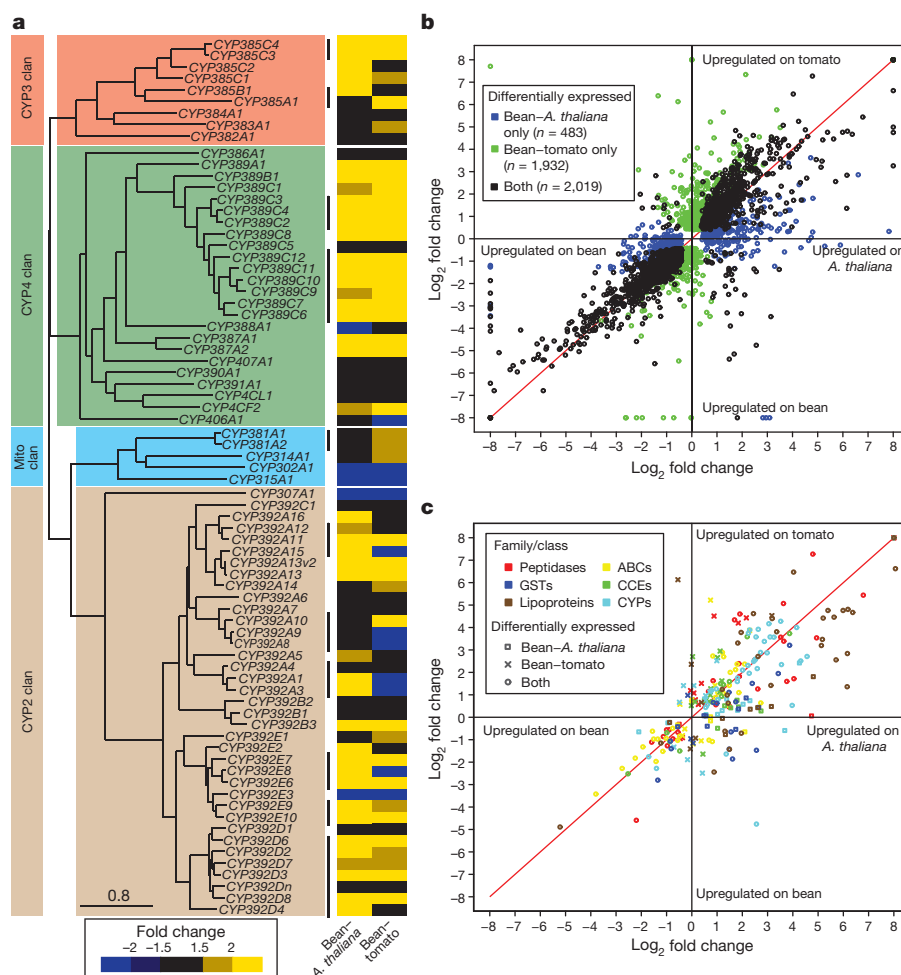
Supplementary Note 5.2) of these genomes showed a gain of about 700 new gene families in the lineage leading to *T. urticae*, plus almost 4,300 genes that are single copy (orphans). More than 1,000 gene families, still present in other arthropods, were lost in *T. urticae*. The 58 gene families that are significantly ( $z$ -score  $>2$ ) expanded in *T. urticae* compared to the other arthropods are shown in Supplementary Note 5.2 and Supplementary Fig. 5.2.1.

## Feeding and detoxification

*Tetranychus urticae* is one of the most striking examples of polyphagy among herbivores and it has an unmatched ability to develop resistance to pesticides<sup>6,10</sup>. We discovered that known gene families implicated in digestion, detoxification and transport of xenobiotics had a unique spider mite composition, and were often expanded when compared to insects (Supplementary Note 6.1). This included a three-fold proliferation of cysteine peptidase genes, particularly C1A papain and the C13 legumain genes (Supplementary Table 6.1.11), consistent with proteolytic digestion based mostly on cysteine peptidase activity<sup>20</sup>. Eighty-six cytochrome P450 (CYP) genes were detected in the *T. urticae* genome, a total number similar to insects but with an expansion of *T. urticae*-specific intronless genes of the CYP2 clan (Supplementary Table 6.1.2). The carboxyl/cholinesterases (CCEs) gene family contained 71 genes, with a single acetylcholinesterase gene (*Ace1*) but two new clades at the root of the neurodevelopmental class of CCEs, representing 34 and 22 CCEs, respectively

(Supplementary Table 6.1.6). A notable case of expansion was found within the family of 32 glutathione *S*-transferases (GSTs) that include a group of 12 Mu-class GSTs that were, until now, believed to be vertebrate-specific (Supplementary Table 6.1.3). Finally, we discovered 39 multidrug resistance proteins belonging to the ATP-binding cassette (ABC) transporters (class C). The repertoire from this class of ABC transporters far exceeds the number (9–14) found in crustaceans, insects, vertebrates and nematodes (Supplementary Table 6.1.8). Few of the genes involved in detoxification had close insect homologues, and only four of the CYP genes could clearly be assigned as orthologues of insect and crustacean CYP genes.

The involvement of these gene families and their spider-mite-specific expansion in host plant adaptation is markedly illustrated by RNA-seq transcriptome profiling of spider mite feeding on its preferred host, bean (*Phaseolus vulgaris*), and on hosts to which the London strain is not adapted: *Arabidopsis thaliana* and tomato (*Solanum lycopersicum*) (Fig. 2) (Supplementary Notes 6.2). We found 24% of all genes to be differentially expressed upon host transfer (Fig. 2a–c); relative to bean, more genes were differentially expressed on tomato than on *A. thaliana* (Supplementary Note 6.2.4 and Supplementary Fig. 6.2.1), but responses were nonetheless correlated (Fig. 2b, c). Genes in the detoxification and peptidase families exhibited the most profound changes (Fig. 2a–c), with expression of nearly half of P450 genes affected by the host plant, including 19 of 39 genes in the intronless CYP392 family and the CYP389 family. These subfamilies



**Figure 2** | Gene expression changes when mites are shifted from *P. vulgaris* (bean) to *A. thaliana* or to *S. lycopersicum* (tomato). **a**, A phylogeny of the cytochrome P450 (CYP) genes and heat map of the response of CYP genes to host transfer. Two-thirds of the genes that are tandemly duplicated or that form clusters (indicated by black vertical lines) are co-regulated. **b**, Global changes in

gene expression after host shift. **c**, Fold changes of important gene family members in digestion and detoxification are colour coded. The analysis of differential expression (**b** and **c**) is with a 5% false discovery rate as assessed with RNA-seq data collected in biological triplicate (fold changes between mean values are plotted).



are spider-mite-specific P450 expansions that define lineage-specific expansions<sup>21</sup>. This finding is unprecedented. In humans, only up to one-third of P450 genes are metabolizing xenobiotics<sup>22</sup>, and in *D. melanogaster* only one-third of the CYP genes are inducible by xenobiotics<sup>23</sup>. The proportion of P450 genes responding to the chemical environment is much greater in the spider mite. Similar patterns were also found within other families (Fig. 2c). For GSTs and CCEs, the expression of Mu and Delta GSTs and the two spider-mite-specific CCE clades were most affected and about one-third of cysteine peptidases, the C1A papains and C13 legumains, were overexpressed after transfer to tomato. More than two-thirds of the CYP and GST genes affected by the host plant are present in clusters of (multiple) tandem duplicated genes. Co-regulation of the majority of tandem duplicates strongly indicates that the ancestral gene was already plant-responsive before duplication, and that a role in plant adaptation may have favoured duplicate retention.

Although these data indicate that spider-mite-specific expansion of known gene families contributes to the ability of spider mites to overcome host defences, many genes differentially regulated upon host transfer lack homology to genes of known function. Notably, among those with the most extreme expression fold-changes are genes that encode putative secreted proteins or lipid-binding proteins. Understanding extracellular binding and transport of small ligands is therefore likely to be important in further dissecting spider mite-plant interactions.

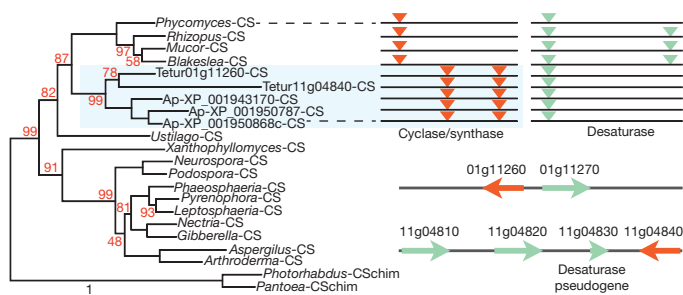
### Lateral gene transfer

Our search for genes related to detoxification and digestion also revealed the existence and surprising expansion of intradiol ring cleavage dioxygenases, genes previously unreported from metazoan genomes but characteristic for bacteria and fungi<sup>24</sup>. We annotated 16 functional genes in this family in *T. urticae*, whereas bacterial genomes usually carry only 1 to 7. They have an average sequence similarity of 43% with the homologue of *Streptomyces avermitilis* and share the conserved 2 His 2 Tyr non-haem iron(III) binding site. These dioxygenases might have evolved to metabolize aromatic compounds found in plant allelochemicals. Other clear instances of lateral gene transfers include (1) the presence of a cobalamin-independent methionine synthase (*MetE*) gene with four predicted introns and up to 58% sequence identity to the *MetE* gene of soil Bacilli (this sequence has not previously been reported in any animal species); (2) two very similar levanase-encoding genes of probable bacterial origin that encode secreted exo-fructosidases upregulated upon feeding on tomato; and (3) a cyanate lyase-encoding gene that might be involved in feeding on cyanogenic plants (Supplementary Table 6.3.1).

We detected two clusters of carotenoid biosynthesis genes in *T. urticae* representing homologues of genes from zygomycete fungi and aphids. The latter are the only animal carotenoid biosynthesis genes known so far, thought to be derived from fungal genes by lateral gene transfer<sup>25</sup>. The unique intron–exon structure of the spider mite and aphid genes and their clustering in phylogenetic analyses is strong evidence that the genes from fungi were transferred only once to arthropods (Fig. 3). The sequence and orientation of the two spider mite clusters indicate that they are the result of an ancient transfer followed by duplications, rearrangements and divergence. They also suggest that a second, more recent transfer occurred between a spider mite and an aphid ancestor, although the sequence of the two transfers remains speculative. Carotenoids are known to have a role in diapause induction in spider mites<sup>26</sup> and our findings indicate that they can also synthesize them.

### Ponasterone A as moulting hormone

Ecdysteroid control of moulting is one of the defining features of arthropods. We detected gene orthologues coding for ecdysteroid biosynthesis enzymes (Supplementary Table 7.1.1)<sup>21</sup>. Surprisingly,



**Figure 3 | Maximum likelihood phylogeny of the fungal and arthropod carotenoid cyclase/synthase (CS) fusion proteins.** The out-group comprises chimaeric assemblies (CSchim) of the closest bacterial sequences of cyclases and synthases. The *T. urticae* and *Acyrtosiphon pisum* sequences form a monophyletic group closely related to the zygomycete sequences. Evidence for a single lateral gene transfer event is also shown by the common intron positions in the cyclase/synthase (orange) and desaturase (green) genes (upper right panel). Two clusters of carotenoid biosynthesis genes are found in *T. urticae*: a tail-to-tail arrangement on scaffold 1 as seen in zygomycetes and aphids, and a more complex head-to-head (re)arrangement on scaffold 11 (bottom right).

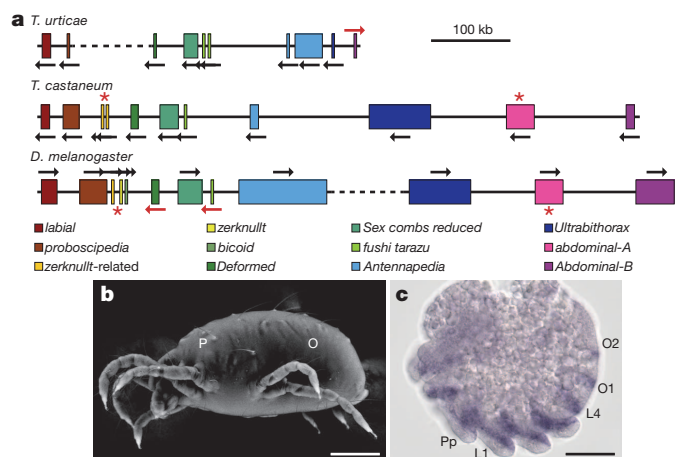
the *T. urticae* genome lacks two P450 genes, *CYP306A1* and *CYP18A1*, encoding, respectively, the biosynthetic C25 hydroxylase and a C26 hydroxylase/oxidase involved in hormone inactivation. The absence of *CYP306A1* indicates that the spider mite uses the ecdysteroid 25-deoxy-20-hydroxyecdysone (ponasterone A) as the moulting hormone, instead of the typical arthropod 20E. This was confirmed by biochemical analysis of spider mite extracts by HPLC–enzyme immunoassay and liquid chromatography/mass spectrometry that identified ponasterone A (Supplementary Note 7). *CYP306A1* and *CYP18A1* form a head-to-head cluster in all insect and crustacean genomes studied so far, therefore their absence from the *T. urticae* genome indicates that they were lost together, affecting both biosynthesis and inactivation pathways of the spider mite moulting hormone. Ponasterone A has been previously identified in some decapod crustaceans, albeit always coincident with 20E (ref. 27), and it is a high potency ligand of all known ecdysteroid receptors.

### Reduced Hox cluster

Hox genes are a conserved set of homeobox-containing transcription factors typically found clustered within the genome and used to establish region-specific identity during early development. The body plan of mites consists of an anterior prosoma and posterior opisthosoma and is further distinguished by an extremely reduced body plan presumably achieved through the fusion of segments (Supplementary Note 8 and Fig. 4b). The ancestral arthropod is predicted to have a Hox cluster with 10 genes<sup>28</sup>. The *T. urticae* genome contains 8 of the canonical 10 genes. The *ftz* gene is present in duplicate, in two closely linked copies; orthologues of *Hox3* and *abdominal A* (*abdA*) were not found (Fig. 4a). This is unusual among chelicerates: all 10 canonical Hox genes are present in the wandering spider<sup>29</sup>. The absence of *abdA* in *T. urticae* correlates with the spider mite's reduced opisthosomal segmentation. Consistent with the absence of *abdA* and a reduced opisthosoma, only two opisthosomal stripes of the segment polarity gene *engrailed* (typically expressed in each arthropod segment) are detected in the developing embryo (Fig. 4c), in contrast to five *engrailed* stripes detected in the opisthosoma of the wandering spider<sup>30</sup>. Although numerous examples correlate morphological variation in arthropods with changes in Hox gene expression, this is the first example that correlates morphological evolution with the loss of a Hox gene within a fully sequenced Hox cluster.

### Nanometre dimensions of *T. urticae* silk

Silk production in spider mites (Fig. 5a, b and Supplementary Note 9) represents a *de novo* evolution of silk-spinning relative to silk production in spiders<sup>7</sup>. Spiders typically spin silk from a complex glandular



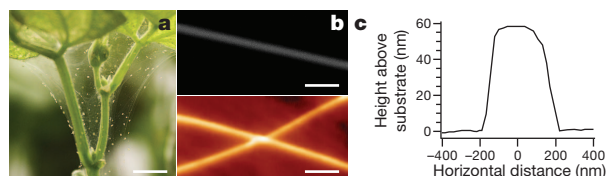
**Figure 4 | Comparative organization of Hox clusters and expression pattern of the *T. urticae* engrailed gene.** **a**, *T. urticae*, *T. castaneum* and *D. melanogaster* Hox clusters. Gene sizes and intergenic distances are shown to scale. Dashed lines represent breaks in the cluster >1 Mb. In *T. urticae*, *fushi tarazu* and *Antennapedia* are present in duplicate whereas *abdominal-A* and *Hox3/zerknullt* are missing (red asterisk). **b**, Variable pressure scanning electron microscopy (SEM) image of adult *T. urticae* with two main body regions indicated: P, prosoma; O, opisthosoma. **c**, *T. urticae* engrailed (*en*) expression pattern. *en* transcripts are detected in five prosomal stripes that correspond to future pedipalpal (Pp), four walking leg (L1–L4) and two opisthosomal (O1 and O2) segments. Scale bars: **b**, 0.125 mm; **c**, 40  $\mu$ m.

abdominal spinneret, whereas *T. urticae* uses paired silk glands connected to the mouth appendages (pedipalps)<sup>31</sup>. Seventeen fibroin genes were uncovered in the genome of *T. urticae* (Supplementary Table 9.1.1) encoding fibroins of unusually high (27–39%) serine content. We performed mechanical testing on fibres deposited by adult and larval mites with an atomic force microscope. This technique measures the Young's modulus of the fibres, which is the ratio of applied stress (tension per cross-sectional area) to the resulting strain (fractional change in length) and describes the stiffness of the material. Young's modulus was higher than or comparable to other natural materials (see Supplementary Table 9.1.2), but *T. urticae* silk fibres are thinner— $54 \pm 3$  nm (adult silk, Fig. 5c) and  $23.3 \pm 0.9$  nm (larval silk), that is, 435–185 times thinner—than the silk fibres of the spider *Nephila clavipes*<sup>32</sup>.

## Concluding remarks

Our analysis of the *T. urticae* genome also included nuclear receptors and neuropeptide genes, immunity-related genes and RNA interference, cuticle protein genes, and DNA methylation (Supplementary Notes 7.3 and 10–12).

The first complete genome of a chelicerate species provides the opportunity for a detailed phylogenomic analysis of arthropods, the most diverse group of animals on Earth. The *T. urticae* genome illustrates the specialized life history of this polyphagous herbivorous pest.



**Figure 5 | *T. urticae* silk structure and dimensions.** **a**, Spider mite colony on a bean plant forming characteristic silk webbing. **b**, SEM image of the spider mite larval silk filament (top), and atomic force microscopy (AFM) image of two larval spider mite silk filaments (bottom). **c**, Height profile of the adult spider mite silk filament obtained from the AFM image. Scale bars: **a**, 0.75 cm; **b**, 1  $\mu$ m.

Striking gene gains include lineage-specific expansions within detoxification gene families and lateral transfer of genes from fungi and bacteria that further expanded in *T. urticae*. The functional significance of these innovations is supported by the upregulation of many of these genes in response to feeding on less preferred host plants.

The genome of the two-spotted spider mite, together with the favourable biological features of the spider mite as a laboratory model including short generation time, easy rearing and tools for gene analysis and gene silencing<sup>33</sup>, provide a novel resource for agriculture that should allow the dissection of pest–plant interactions and development of alternative tools for plant protection. Finally, evolutionary innovation in the process of *T. urticae* silk production expands the repertoire of potential chelicerate biomaterials (such as the well-known spider silk) with a natural biomaterial at the nanometre scale.

## METHODS SUMMARY

All genomic sequencing reads were collected with standard Sanger sequencing protocols. RNA sequencing was performed with Illumina RNA-seq protocols. Annotation of the *T. urticae* genome was done using the gene prediction platform EuGene. The complete genome annotation is available at <http://bioinformatics.psb.ugent.be/webtools/bogas/>. The *T. urticae* (London) genome project was registered under the INSDC project ID 71041.

Received 30 June; accepted 17 October 2011.

- Edgecombe, G. D. Arthropod phylogeny: An overview from the perspectives of morphology, molecular data and the fossil record. *Arthropod Struct. Dev.* **39**, 74–87 (2010).
- Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
- Dunlop, J. A. & Selden, P. A. Calibrating the chelicerate clock: a paleontological reply to Jayaprakash and Hoy. *Exp. Appl. Acarol.* **48**, 183–197 (2009).
- Dunlop, J. A. Geological history and phylogeny of Chelicerata. *Arthropod Struct. Dev.* **39**, 124–142 (2010).
- Walter, D. E. & Proctor, H. C. *Mites: Ecology, Evolution and Behaviour* (CABI Publishing, 1999).
- Jeppson, L. R., Keifer, H. H. & Baker, E. W. *Mites Injurious to Economic Plants* (Univ. California Press, 1975).
- Gerson, U. in *Spider Mites: their Biology, Natural Enemies and Control* Vol. 1A (eds Helle, W. & Sabelis, M. W.) 223–232 (Elsevier, 1985).
- Boubou, A., Migeon, A., Roderick, G. K. & Navajas, M. Recent emergence and worldwide spread of the red tomato spider mite, *Tetranychus evansi*: genetic variation and multiple cryptic invasions. *Biol. Invasions* **13**, 81–92 (2011).
- Migeon, A. *et al.* Modelling the potential distribution of the invasive tomato red spider mite, *Tetranychus evansi* (Acari: Tetranychidae). *Exp. Appl. Acarol.* **48**, 199–212 (2009).
- Van Leeuwen, T., Vontas, J., Tsagkakarakou, A., Dermauw, W. & Tirry, L. Acaricide resistance mechanisms in the two-spotted spider mite *Tetranychus urticae* and other important Acari: a review. *Insect Biochem. Mol. Biol.* **40**, 563–572 (2010).
- Khajehali, J., Van Nieuwenhuysse, P., Demaeght, P., Tirry, L. & Van Leeuwen, T. Acaricide resistance and resistance mechanisms in *Tetranychus urticae* populations from rose greenhouses in the Netherlands. *Pest Manag. Sci.* **67**, 1424–1433 (2011).
- BOGAS. Bioinformatics Gent Online Genome Annotation Service (<http://bioinformatics.psb.ugent.be/webtools/bogas/>) (2011).
- VectorBase. *Ixodes scapularis* Wikel annotation, IscaW1 (<http://iscapularis.vectorbase.org/>) (2008).
- Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res.* **19**, 1776–1785 (2009).
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
- Richards, S. *et al.* The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949–955 (2008).
- Oliver, J. H. Cytogenetics of mites and ticks. *Annu. Rev. Entomol.* **22**, 407–429 (1977).
- Kirkness, E. F. *et al.* Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl Acad. Sci. USA* **107**, 12168–12173 (2010).
- Cornman, S. R. *et al.* Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. *BMC Genom.* **11**, 602 (2010).
- Carrillo, L. *et al.* Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. *Plant Cell Rep.* **30**, 101–112 (2011).
- Feyereisen, R. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochim. Biophys. Acta* **1814**, 19–28 (2011).
- Guengerich, F. P., Wu, Z. L. & Bartleson, C. J. Function of human cytochrome P450s: characterization of the orphans. *Biochem. Biophys. Res. Commun.* **338**, 465–469 (2005).



# Neuronal filtering of multiplexed odour representations

Francisca Blumhagen<sup>1†\*</sup>, Peixin Zhu<sup>1\*</sup>, Jennifer Shum<sup>1†\*</sup>, Yan-Ping Zhang Schärer<sup>1\*</sup>, Emre Yaksi<sup>2</sup>, Karl Deisseroth<sup>3</sup> & Rainer W. Friedrich<sup>1,4</sup>

**Neuronal activity patterns contain information in their temporal structure, indicating that information transfer between neurons may be optimized by temporal filtering. In the zebrafish olfactory bulb, subsets of output neurons (mitral cells) engage in synchronized oscillations during odour responses, but information about odour identity is contained mostly in non-oscillatory firing rate patterns. Using optogenetic manipulations and odour stimulation, we found that firing rate responses of neurons in the posterior zone of the dorsal telencephalon (Dp), a target area homologous to olfactory cortex, were largely insensitive to oscillatory synchrony of mitral cells because passive membrane properties and synaptic currents act as low-pass filters. Nevertheless, synchrony influenced spike timing. Moreover, Dp neurons responded primarily during the decorrelated steady state of mitral cell activity patterns. Temporal filtering therefore tunes Dp neurons to components of mitral cell activity patterns that are particularly informative about precise odour identity. These results demonstrate how temporal filtering can extract specific information from multiplexed neuronal codes.**

The temporal structure of neuronal activity patterns contains information that may be retrieved by temporal filtering operations<sup>1–8</sup>. Oscillatory synchronization has been implicated in sensory and cognitive processing based on the assumption that synchronization enhances the impact of neuronal ensembles on their targets<sup>3,5,7,8</sup>. Indeed, coincidence detection is a generic neuronal filtering operation that can be further supported by active conductances and synaptic microcircuits. However, few studies have directly demonstrated the impact of synchrony on neuronal responses in the intact brain<sup>9–12</sup>. Moreover, information can also be encoded by other temporal features of population activity patterns<sup>6,13–20</sup>, and the information transmitted during a sensory response can change over time<sup>20–24</sup>. The retrieval of information from neuronal activity patterns thus depends on temporal filters that tune higher-order neurons to specific temporal structure in their input. Filtering is particularly important when different temporal features simultaneously transmit distinct messages (multiplexing)<sup>13,15,24</sup>.

We explored temporal filtering during transformations of spatio-temporal activity patterns in the olfactory system of zebrafish. Odours stimulate combinations of mitral cells in the olfactory bulb, which project to multiple higher brain areas. In zebrafish, mitral cell activity patterns are dynamically reorganized during the initial phase of an odour response before they approach a steady state<sup>5,22,24,25</sup>. After approximately 400 ms, two important consequences of this reorganization emerge (Supplementary Fig. 1)<sup>24</sup>. First, odour-dependent subsets of mitral cells rhythmically synchronize their action potentials at a frequency near 20 Hz. Activity patterns across synchronized mitral cells are highly correlated in response to similar odours and contain information about molecular categories, but they are not well suited for precise odour identification. Second, patterns of non-synchronized spikes become decorrelated and informative about precise odour identity. Mitral cell activity patterns in the steady state therefore simultaneously convey information about complementary stimulus features<sup>24</sup>. Extracting information from these multiplexed activity patterns

depends critically on temporal filtering by their targets. Precise odour identity could be efficiently decoded from late activity patterns across non-synchronized mitral cells, but this may be complicated by synchronized mitral cell ensembles as synchronization can have a strong impact on target neurons.

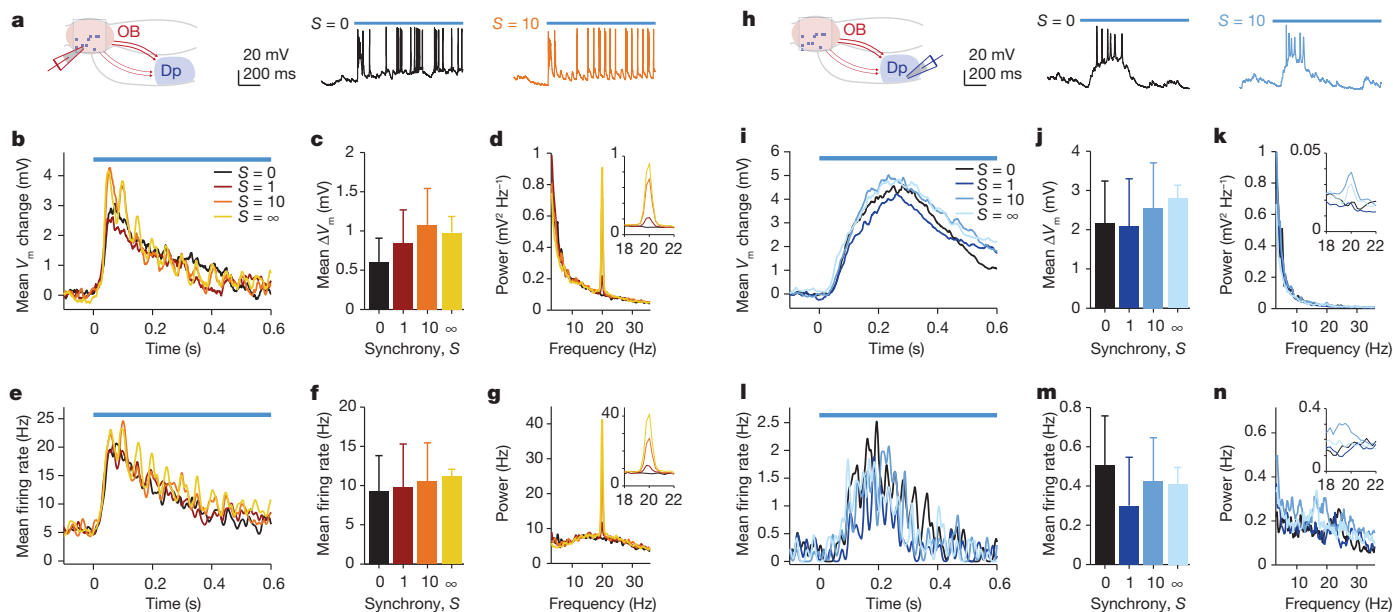
## Optogenetic manipulation of synchrony

We first used an optogenetic approach to examine the impact of synchrony among mitral cells on neuronal responses in the posterior zone of the dorsal telencephalon (Dp), a major target of the olfactory bulb that is homologous to olfactory cortex<sup>26,27</sup>. In transgenic zebrafish expressing channelrhodopsin-2 (Chr2)<sup>28,29</sup> in sensory afferents<sup>30</sup>, patterns of blue laser light were projected onto the ventral olfactory bulb using a digital micromirror device (DMD). Each of the  $1,024 \times 768$  mirrors in the DMD was controlled independently with submillisecond temporal resolution. Light patterns consisted of distributed active squares ( $50 \times 50 \mu\text{m}^2$  or  $20 \times 20 \mu\text{m}^2$ ; Fig. 1a and Supplementary Fig. 2). Activation of individual squares near a mitral cell for 5 ms evoked depolarizations up to 10 mV and action potential firing (Supplementary Fig. 3). In a stimulus pattern, each mirror within an active square was switched on for 5 ms at a mean event rate of 20 Hz. Synchrony among pixels was then varied by periodically modulating the probability of events at a common carrier frequency of 20 Hz (Supplementary Fig. 4 and Supplementary Movies 1 and 2). Synchrony was quantified by a synchronization index,  $S$ , that ranges between zero (independent Poisson-distributed events) and infinity (perfectly synchronous events; Supplementary Fig. 4).

Mitral cells responded robustly to optical stimulation. As  $S$  was increased ( $S = 0, 1, 10, \infty$ ), 20 Hz oscillatory fluctuations emerged in the membrane potential of individual neurons (Fig. 1a), as well as in averaged recordings (Fig. 1b–d;  $n = 161$  responses of 49 mitral cells from 13 fish at each  $S$ ;  $3.4 \pm 1.1$  trials per stimulus; mean  $\pm$  s.d.) and in the local field potential (LFP; Fig. 2 and Supplementary Fig. 5). The overall magnitude and time course of the mitral cell response,

<sup>1</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. <sup>2</sup>Neuroelectronics Research Flanders, Kapeldreef 75, 3001 Leuven, Belgium. <sup>3</sup>Department of Bioengineering, Stanford University, Stanford, California 94305, USA. <sup>4</sup>University of Basel, 4003 Basel, Switzerland. <sup>†</sup>Present addresses: Novartis Pharma GmbH, 90327 Nuremberg, Germany (F.B.); Department of Neurology and Neurological Sciences, Stanford University, Stanford, California 94305, USA (J.S.).

\*These authors contributed equally to this work.



**Figure 1 | Optogenetic manipulation of mitral cell synchrony and its effect on Dp neurons.** **a**, Schematic illustration of optical stimulation in the olfactory bulb (OB) and responses of a mitral cell to the same spatial pattern at two different synchronies  $S$ . **b**, Mean membrane potential ( $V_m$ ) response of mitral cells to the same optical stimulus patterns presented at four different synchronies  $S$  ( $n = 161$  stimuli in 49 mitral cells at each  $S$ ). Action potentials were removed before averaging by median filtering. **c**, Mean membrane potential change during the first second of optical stimulation as a function of synchrony. **d**, Power spectrum of the membrane potential, averaged over trials

and neuron–stimulus pairs, during the first second of the response to optical stimulus patterns with different synchrony. Inset shows an enlargement around 20 Hz. **e–g**, Time course of mean mitral cell firing rate, mean mitral cell firing rate during the first second, and average power spectrum of mitral cell firing rate in response to patterns with different synchrony. Continuous firing rate traces were generated by convolving action potential events with a Gaussian kernel (s.d. = 5 ms). **h–n**, Same analysis for responses of Dp neurons. Error bars show s.d.

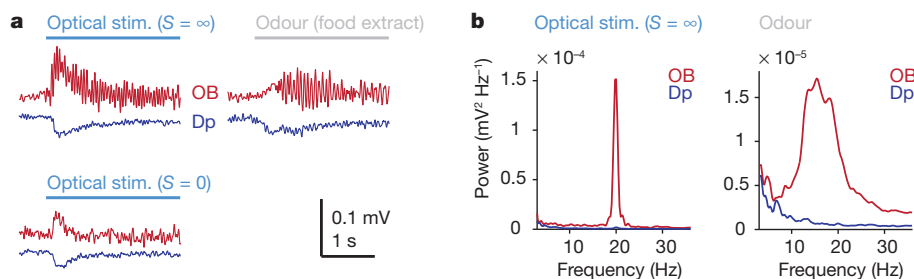
however, remained nearly constant (Fig. 1b, c). Similarly, increasing  $S$  strongly enhanced oscillatory synchronization of firing rate responses, with minimal effects on the magnitude and overall time course (Fig. 1e–g). Optical stimulation therefore allowed us to selectively modulate oscillatory synchrony among mitral cells.

We then stimulated the olfactory bulb with the same light patterns while recording from different types of Dp neurons (Fig. 1h and Supplementary Fig. 6). Dp neurons had low spontaneous activity ( $0.1 \pm 0.5$  Hz; mean  $\pm$  s.d.) and usually responded to optical stimulation with a depolarization, but only a minority of responses were suprathreshold (77 out of 372 neuron–stimulus pairs; 21%). The mean membrane potential change and overall response time course were not substantially affected by synchrony (Fig. 1i, j;  $n = 94$  responses of 34 Dp neurons from 10 fish at each  $S$ ). Unlike in the olfactory bulb, increasing synchrony did not produce prominent oscillations in the membrane potential (Fig. 1h, i) or in the LFP measured in Dp (Fig. 2 and Supplementary Fig. 5). Even when depolarizations were large and when synchrony was high, oscillatory

membrane potential fluctuations were small and occurred on top of a large, slow depolarization (Supplementary Fig. 7). Power spectral analysis revealed only a slight increase in membrane potential fluctuations at 20 Hz (Fig. 1k; note different scaling of inset). Action potential firing of Dp neurons did also not systematically depend on synchrony and showed no prominent 20 Hz oscillation at any level of synchronization (Fig. 1l–n). Moreover, switching synchrony from  $S = 0$  to  $S = \infty$  during optical stimulation failed to evoke an obvious response (Supplementary Fig. 8). Nevertheless, spikes evoked by oscillatory input occurred around a preferred phase of the LFP oscillation in the olfactory bulb (Supplementary Fig. 9). Oscillatory synchrony can therefore influence the timing of action potentials in Dp, but it did not increase the overall depolarization and firing rates.

### Impact of synchrony on odour responses

To examine the impact of synchrony on odour responses of Dp neurons we performed whole-cell recordings during stimulation with natural odours (food extracts and amino acid odours;  $n = 166$  neuron–odour



**Figure 2 | Local field potential responses.** **a**, Left: simultaneously recorded LFP responses in the olfactory bulb (OB; red) and Dp (blue) to optical stimulation with high synchrony ( $S = \infty$ ; top) and no synchrony ( $S = 0$ ; bottom). Right: LFP responses to odour stimulation (food extract; 5–40 Hz). **b**, Power spectra of LFP responses to optical stimulation with high synchrony

(left;  $S = \infty$ ) and odour stimulation (right) in the olfactory bulb (red) and in Dp (blue). Power spectra were calculated for each trial and averaged over trials and stimuli (olfactory bulb, odour:  $n = 208$  stimuli in 17 fish; Dp, odour:  $n = 20$  stimuli in 4 fish; olfactory bulb and Dp, optical:  $n = 16$  stimuli in 4 fish).



pairs, 57 Dp neurons in 43 fish; mean number of repetitions per stimulus,  $4.1 \pm 1.8$ ). As observed previously<sup>26</sup>, responses of Dp neurons were often depolarizing, but many remained subthreshold (Supplementary Fig. 10). Power spectral analysis showed no evidence for oscillatory modulation of the membrane potential, spiking activity, or LFP in Dp during odour stimulation (Fig. 2 and Supplementary Fig. 5).

In 122 neuron–odour pairs (492 individual trials), we simultaneously recorded the LFP oscillation in the olfactory bulb and compared it to membrane potential fluctuations of Dp neurons in the same frequency band (5–40 Hz; Fig. 3a). In some cases, a weak oscillatory cross-correlation was observed (Supplementary Fig. 11); however, membrane potential fluctuations were substantially smaller than the slow depolarization. On average, the amplitude of membrane potential fluctuations (filtered 5–40 Hz) was  $0.7 \pm 0.7$  mV (mean  $\pm$  s.d.; mean maximal fluctuations,  $2.4 \pm 1.9$  mV). In the same neurons, the mean difference between the resting potential and action potential threshold was  $27.7 \pm 5.9$  mV (Fig. 3b). The amplitude of the slow depolarization had a mean of  $15.4 \pm 12.7$  mV ( $\pm$  s.d.) and varied substantially across neurons and odours (Fig. 3b). Spike output therefore depended primarily on the magnitude of the slow depolarization, rather than on oscillatory fluctuations. Nevertheless, spikes of Dp neurons occurred around a preferred phase of the LFP oscillation in the olfactory bulb (Supplementary Fig. 9), indicating that oscillatory synchronization influenced action potential timing.

To confirm further that oscillatory fluctuations are not the primary drive underlying action potentials, we compared subthreshold and suprathreshold responses of the same Dp neurons to different odours (Fig. 3c;  $n = 13$  Dp neurons;  $6.7 \pm 5.5$  and  $6.7 \pm 3.8$  subthreshold and

suprathreshold trials per neuron, respectively). The magnitude of the slow component was, on average, 10.1 mV larger in suprathreshold responses than in subthreshold responses ( $15.5 \pm 10.7$  mV versus  $25.6 \pm 12.0$  mV;  $P < 10^{-5}$ ), whereas the mean amplitude of fluctuations in the 5–40 Hz band was only 0.3 mV larger ( $0.8 \pm 0.5$  mV versus  $1.1 \pm 0.5$  mV;  $P < 10^{-4}$ ; Fig. 3c). Hence, oscillatory fluctuations did not substantially contribute to the overall depolarization underlying action potential firing.

### Low-pass filtering mechanisms

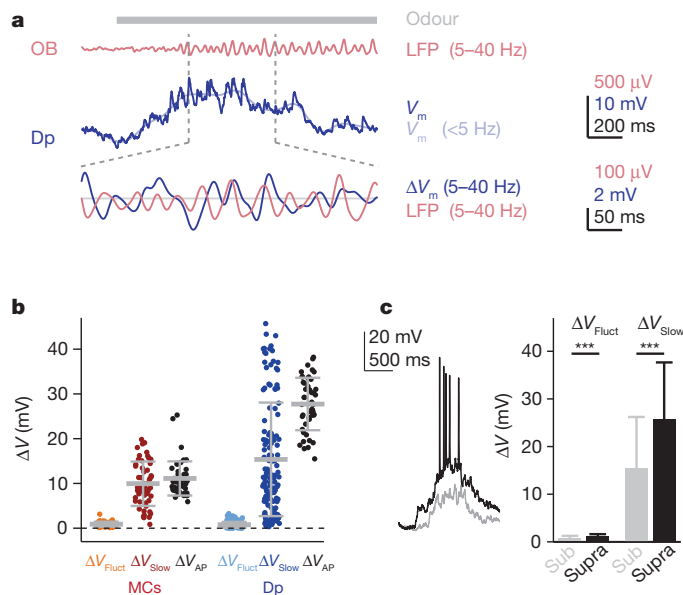
To explore why synchrony had no detectable effect on firing rates of Dp neurons we first measured their membrane time constants by step current injections. In randomly selected neurons, time constants were  $29 \pm 12$  ms (mean  $\pm$  s.d.;  $n = 27$  neurons) and  $45 \pm 18$  ms ( $n = 30$  neurons) in response to hyperpolarizing and depolarizing steps, respectively (Fig. 4a). In neurons expressing GFP from the promoter of the vesicular glutamate transporter 2a, a marker for glutamatergic neurons, time constants were  $67 \pm 43$  ms and  $79 \pm 24$  ms after hyperpolarizing and depolarizing pulses, respectively ( $n = 14$  each; GFP-negative neurons,  $33 \pm 13$  ms and  $41 \pm 14$  ms,  $n = 5$  each;  $P < 0.01$ ; Fig. 4a). These time constants are long compared to the temporal precision of spike synchronization in the olfactory bulb<sup>24</sup> and to the period of the oscillation. Passive membrane properties of Dp neurons therefore support low-pass filtering rather than coincidence detection, particularly in glutamatergic neurons.

We then injected sinusoidal currents with fixed or modulated frequencies to examine whether oscillatory input at frequencies near 20 Hz is amplified by active conductances (Supplementary Fig. 12). However, the voltage response decreased gradually as a function of frequency (Fig. 4b), indicating passive low-pass filtering<sup>2</sup>. To examine responses in the spiking regime we simulated synaptic inputs of 20 neurons, each firing at a mean rate of 20 Hz, and varied their oscillatory synchrony by varying  $S$  (Supplementary Fig. 4). Simulated spike patterns (10-s duration) were convolved with idealized excitatory post-synaptic current waveforms and injected into Dp neurons in current clamp. The amplitude was adjusted to evoke action potential firing at  $<10$  Hz for  $S = 0$ . On average, firing rates increased slightly with synchrony, but this effect was not significant (Fig. 4c, d;  $P > 0.3$  for all pairwise comparisons). Together, these results indicate that Dp neurons are not intrinsically tuned to detect synchronized input near the natural oscillation frequency but act as passive low-pass filters, unlike Kenyon cells in the mushroom body of insects<sup>9,10,31</sup>.

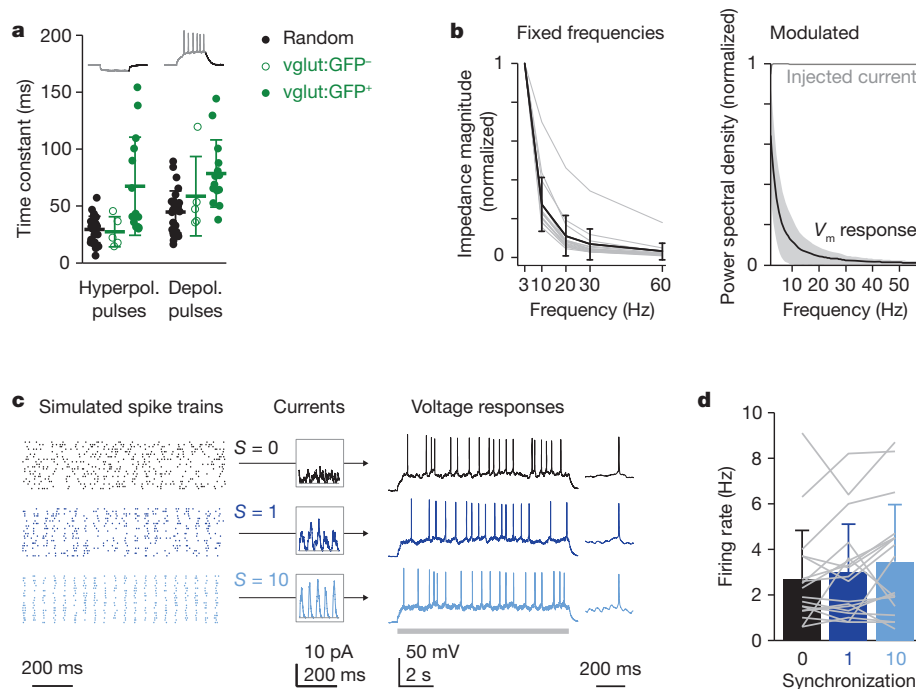
In many brain areas, a delay between excitatory and inhibitory input can establish a narrow time window for the integration of coincident synaptic input<sup>32</sup>. This mechanism selects synchronized input in the insect mushroom body<sup>10</sup> and has been proposed also for olfactory cortex<sup>33–35</sup>. To examine whether delayed inhibition supports coincidence detection in Dp neurons, we recorded excitatory and inhibitory synaptic currents evoked by 20 Hz electrical stimulation of the olfactory tract and analysed synaptic integration in a conductance-based model neuron (Supplementary Fig. 13). Although inhibitory currents were delayed, oscillatory synchrony was not efficiently detected by the model neuron because passive membrane properties act as strong low-pass filters at 20 Hz, and because inhibitory synaptic currents were slow (Supplementary Fig. 13). Our results therefore indicate that slow passive properties, the absence of active amplification, and the slow time course of collective synaptic currents can account, at least in part, for the insensitivity of Dp neurons to synchrony in their inputs.

### Late responses of Dp neurons

When responses to optical stimulation were averaged over all neurons, stimuli and synchronizations, the peak of the membrane potential and firing rate responses in Dp ( $n = 372$  responses from 34 Dp neurons) occurred approximately 200 ms after the peak of mitral cell responses ( $n = 644$  responses from 49 mitral cells; Fig. 5a, b). This effect was



**Figure 3 | Spiking responses to odours in Dp are driven by slow depolarization.** **a**, Simultaneous recording of LFP in the olfactory bulb (5–40 Hz; red) and membrane potential of a Dp neuron (dark blue) during odour stimulation (grey bar). Light-blue trace shows low-pass (5 Hz) filtered membrane potential. Bottom: comparison of membrane potential fluctuations (difference between original and low-pass filtered membrane potential) and LFP after band-pass-filtering (5–40 Hz) in a 500-ms time window. **b**, Amplitude of membrane potential fluctuations in the oscillatory frequency band (5–40 Hz;  $\Delta V_{Fluct}$ ), amplitude of the slow depolarization ( $\Delta V_{Slow}$ ) in the same trials, and difference between resting potential and spike threshold ( $\Delta V_{AP}$ ) in mitral cells (MCs; left) and Dp neurons (right). Horizontal lines and error bars show mean  $\pm$  s.d. **c**, Left: subthreshold and suprathreshold responses of a Dp neuron to different odours (Arg, 100  $\mu$ M, and mixture of Arg and Tyr, 100  $\mu$ M each). Right: mean amplitude of membrane potential fluctuations ( $\pm$  s.d.) in the oscillatory frequency band (5–40 Hz;  $\Delta V_{Fluct}$ ) and the slow depolarization ( $\Delta V_{Slow}$ ) in subthreshold and suprathreshold responses of the same Dp neurons ( $n = 13$ ). \*\*\* $P < 10^{-4}$ .



**Figure 4 | Intrinsic properties of Dp neurons.** **a**, Membrane time constant measured by exponential fits to the membrane potential decay after hyperpolarizing (left) and depolarizing (right) current injection (500 ms). **b**, Normalized impedance of Dp neurons measured by injection of sine currents at five fixed frequencies (left;  $n = 15$ ), and power spectral density of the membrane potential response to injection of frequency-modulated sine currents (right; 1–60 Hz;  $n = 9$ ). Grey traces (left) represent data from individual neurons; grey shading (right) shows s.d. Black traces show mean; error bars show s.d. **c**, Left: examples of simulated spike train patterns. Each row

in a spike train pattern shows action potential times of one simulated neuron. Centre: segments of summed currents from spike train patterns with different synchrony. Horizontal grey lines represent zero current. Right: responses of a Dp neuron to currents from spike train patterns with different synchrony. The membrane potential evolution around a single action potential is shown on the right. The grey bar shows time of current injection. **d**, Mean firing frequency ( $\pm$  s.d.) of Dp neurons upon injection of currents from spike train patterns with different synchrony. Grey lines show results from individual cells.

independent of synchrony (compare Figs 1b, e and Figs 1i, l) and also observed when only strong responses ( $\geq 10$  s.d. of fluctuations before stimulus onset) were considered (Fig. 5c, d). In response to odours, mitral cell firing rates changed during the initial phase of the response and stabilized after approximately 400 ms ( $n = 52$  mitral cell–odour pairs; 12 mitral cells in 8 fish), as observed previously<sup>22,24</sup>. Responses of Dp neurons evolved more slowly and peaked substantially later, even when only strong depolarizing responses ( $\geq 10$  s.d. of pre-odour fluctuations) were considered (Fig. 5e–h and Supplementary Fig. 10). Peak firing of Dp neurons occurred approximately 500 ms later than in mitral cells. Hence, responses of many Dp neurons occurred during the decorrelated steady state of mitral cell activity patterns.

Filtering of oscillatory input and late responses of Dp neurons should favour the detection of decorrelated mitral cell activity patterns. To examine this we measured population activity patterns evoked by three groups of similar amino acid odours (10  $\mu$ M) using temporally deconvolved two-photon calcium imaging<sup>36</sup>. Odours within each group activate overlapping combinations of glomeruli<sup>37</sup> and evoke mitral cell activity patterns that become decorrelated<sup>22,24,38</sup>. In Dp, correlations between activity patterns were already lower than in the olfactory bulb during the initial phase of the odour response<sup>22,24,38</sup> and subsequently decreased further (Fig. 5i and Supplementary Fig. 14). The transformation of activity patterns between the olfactory bulb and Dp therefore results in decorrelated odour representations, consistent with the temporal filtering properties of Dp neurons.

## Discussion

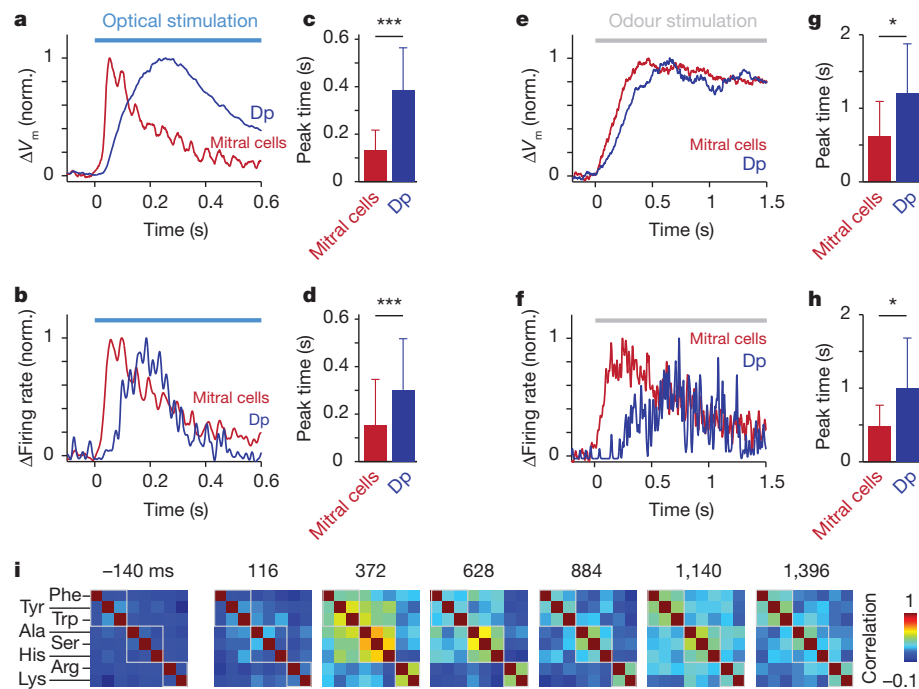
We found that the response intensity of neurons in Dp, a cortex-like target of the olfactory bulb in zebrafish, is largely insensitive to 20 Hz oscillatory synchrony in its input, demonstrating that synchrony does not necessarily enhance responses of target neurons. The impact of synchrony is attenuated because passive properties of Dp neurons and

synaptic currents act as low-pass filters. As a consequence, Dp neurons do not directly decode the information contained in the oscillatory synchrony of mitral cell ensembles. However, this information may not be lost because mitral cell synchrony is still reflected in the timing of action potentials in Dp. Moreover, information contained in the synchrony of mitral cell ensembles may be processed by other target areas of the olfactory bulb. Indeed, odour-evoked LFP oscillations phase-locked to those in the olfactory bulb were detected in a distinct area in the central telencephalon (Supplementary Fig. 15).

Many suprathreshold responses of Dp neurons occurred after the peak of the mitral cell response, when mitral cell activity patterns approached their steady state. Multiple mechanisms may contribute to this delay. First, the long membrane time constants of Dp neurons indicate that the integration of synaptic inputs is slow. Second, results from mammalian piriform cortex demonstrated that feed-forward inhibitory neurons are more sensitive to mitral cell input and more broadly tuned than principal neurons<sup>27,34,39,40</sup>. Feed-forward inhibition by a small subset of interneurons may thus suppress responses of other neurons during the early phase of an odour response. Third, inhibitory input to piriform cortex neurons undergoes depression whereas excitatory input undergoes facilitation in response to trains of electrical stimulation in brain slices<sup>39</sup>. These observations are consistent with synaptic currents evoked by olfactory tract stimulation in Dp neurons (Y.-P.Z.S. and R.W.F., unpublished data) and suggest that the net excitatory input increases during later response phases. Fourth, in a recurrent network, the summation of direct and recurrent excitatory input in individual neurons is expected to be maximally effective in the steady state when both inputs are stable. Responses of Dp neurons may therefore be low initially because the network is unable to follow dynamic mitral cell input.

Temporal filtering attenuates the impact of synchronized mitral cell ensembles and early mitral cell activity patterns, which contain





**Figure 5 | Late responses of Dp neurons.** **a**, Time course of membrane potential responses averaged over all optical stimulus patterns and synchronies (normalized to maximum). Red, mitral cells ( $n = 644$  stimulus–neuron pairs in 49 mitral cells); blue, Dp neurons ( $n = 372$  stimulus–neuron pairs in 34 Dp neurons). **b**, Average time course of firing rate responses (normalized; same responses). **c**, **d**, Mean time to peak depolarization ( $\pm$  s.d.) of strong membrane potential responses (**c**;  $\geq 10$  s.d. of baseline fluctuations; mitral cells,  $n = 25$  stimulus–neuron pairs; Dp neurons,  $n = 82$ ) and strong firing rate changes (**d**; mitral cells,  $n = 137$ ; Dp neurons,  $n = 58$ ). **e**, Mean membrane potential time course (normalized) of depolarizing odour responses. Mitral cells and Dp neurons were recorded in the same fish or under the same conditions (mitral

cells,  $n = 34$  neuron–odour pairs; Dp neurons,  $n = 33$ ). **f**, Mean time course of odour-evoked firing rate responses (normalized). **g**, **h**, Mean time-to-peak ( $\pm$  s.d.) of strong depolarizing responses (**g**;  $\geq 10$  s.d.; mitral cells,  $n = 12$ ; Dp neurons,  $n = 14$ ) and strong firing rate responses to odours (**h**;  $\geq 10$  s.d.; mitral cells,  $n = 15$ ; Dp neurons,  $n = 19$ ).  $*P < 0.05$ ;  $***P < 10^{-9}$ . **i**, Pearson correlation between activity patterns ( $n = 405$  neurons from 5 fish) evoked by different odours as a function of time. Activity patterns were measured using temporally deconvolved two-photon calcium imaging. Time relative to stimulus onset ( $t = 0$ ) is shown above each correlation matrix in milliseconds. Left correlation matrix ( $t = -140$  ms) shows pre-stimulus correlations.

information about molecular categories but are not well suited for fine odour discrimination. Temporal filtering therefore tunes Dp neurons to features of mitral cell responses that are particularly informative about precise odour identity. Consistent with these filtering operations, activity patterns evoked by similar odours were more distinct in Dp than in the olfactory bulb. The transformation between the olfactory bulb and Dp therefore establishes specific and decorrelated odour representations, which is a prerequisite for higher-order computations such as pattern storage by auto-associative networks.

Contrary to neurons in Dp, Kenyon cells in the insect mushroom body are exquisitely sensitive to oscillatory synchronization<sup>10,41</sup> and respond preferentially during the initial phase of an odour response<sup>9,25</sup>. Pattern read out in the mushroom body is therefore different from Dp. However, odour representations differ already at the previous processing stage: in the antennal lobe, response patterns are dense, synchronized ensembles represent odour identity, and the dynamic phase of the odour response is more informative than the steady state<sup>6,25</sup>. Moreover, the convergence onto second-order neurons in the insect mushroom body is high<sup>42</sup>, unlike in olfactory cortex<sup>27</sup>. Under these conditions, the detection of synchronized ensembles results in highly selective, robust and sparse odour representations<sup>9,10,41</sup>. Second-order processing therefore extracts information about precise odour identity both in zebrafish and insects. In zebrafish, temporal constraints on odour processing may be less strict than in other species because zebrafish live in non-turbulent environments and do not sniff<sup>23,44</sup>. As a consequence, zebrafish may afford to use steady-state activity patterns for neuronal computations. In general, our results show that temporal filters can be tuned to extract specific information from multiplexed neuronal codes, indicating that

multiplexing can be used by the brain to efficiently encode and transmit information.

## METHODS SUMMARY

Experiments were performed in an explant preparation of the intact brain and nose of adult zebrafish<sup>22</sup>. Odour stimulation, electrophysiological recordings and temporally deconvolved calcium imaging were performed as described<sup>12,26,36</sup>. Chr2 was expressed in sensory afferents under the control of the Tet system as described<sup>30</sup> or, in a few fish, using the promoter of olfactory marker protein<sup>45</sup>. Spatio-temporal patterns of blue laser light were projected onto the olfactory bulb through the objective of an upright microscope using a DMD with an accelerator board (DMD Discovery 1100 with ALP-1; Texas Instruments) that enabled temporal resolution up to 0.125 ms.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 17 May; accepted 13 October 2011.

Published online 13 November 2011.

- Gerstner, W., Kreiter, A. K., Markram, H. & Herz, A. V. Neural codes: firing rates and beyond. *Proc. Natl Acad. Sci. USA* **94**, 12740–12741 (1997).
- Hutcheon, B. & Yarom, Y. Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci.* **23**, 216–222 (2000).
- Fries, P. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
- Gütig, R. & Sompolinsky, H. The tempotron: a neuron that learns spike timing-based decisions. *Nature Neurosci.* **9**, 420–428 (2006).
- Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nature Rev. Neurosci.* **3**, 884–895 (2002).
- Rabinovich, M., Huerta, R. & Laurent, G. Transient dynamics for neural processing. *Science* **321**, 48–50 (2008).
- Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).

8. Singer, W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* **24**, 49–65 (1999).
9. Perez-Orive, J., Bazhenov, M. & Laurent, G. Intrinsic and circuit properties favor coincidence detection for decoding oscillatory input. *J. Neurosci.* **24**, 6037–6047 (2004).
10. Perez-Orive, J. *et al.* Oscillations and sparsening of odor representations in the mushroom body. *Science* **297**, 359–365 (2002).
11. Azouz, R. & Gray, C. M. Adaptive coincidence detection and dynamic gain control in visual cortical neurons *in vivo*. *Neuron* **37**, 513–523 (2003).
12. Bruno, R. M. & Sakmann, B. Cortex is driven by weak but synchronously active thalamocortical synapses. *Science* **312**, 1622–1627 (2006).
13. Gawne, T. J., Kjaer, T. W. & Richmond, B. J. Latency: another potential code for feature binding in striate cortex. *J. Neurophysiol.* **76**, 1356–1360 (1996).
14. Hopfield, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
15. Golisch, T. & Meister, M. Rapid neural coding in the retina with relative spike latencies. *Science* **319**, 1108–1111 (2008).
16. Junek, S., Kludt, E., Wolf, F. & Schild, D. Olfactory Coding with Patterns of Response Latencies. *Neuron* **67**, 872–884 (2010).
17. Spors, H., Wachowiak, M., Cohen, L. B. & Friedrich, R. W. Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *J. Neurosci.* **26**, 1247–1259 (2006).
18. Bathellier, B., Buhl, D. L., Accolla, R. & Carleton, A. Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. *Neuron* **57**, 586–598 (2008).
19. Cury, K. M. & Uchida, N. Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* **68**, 570–585 (2010).
20. Maass, W., Natschlager, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
21. Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* **400**, 869–873 (1999).
22. Friedrich, R. W. & Laurent, G. Dynamic optimization of odor representations in the olfactory bulb by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).
23. Ringach, D. L., Hawken, M. J. & Shapley, R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* **387**, 281–284 (1997).
24. Friedrich, R. W., Habermann, C. J. & Laurent, G. Multiplexing using synchrony in the zebrafish olfactory bulb. *Nature Neurosci.* **7**, 862–871 (2004).
25. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).
26. Yaksi, E., von Saint Paul, F., Niessing, J., Bunschuh, S. T. & Friedrich, R. W. Transformation of odor representations in target areas of the olfactory bulb. *Nature Neurosci.* **12**, 474–482 (2009).
27. Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
28. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
29. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
30. Zhu, P. *et al.* Optogenetic dissection of neuronal circuits in zebrafish using viral gene transfer and the Tet system. *Front. Neural Circuits* **3**, 21 (2009).
31. Laurent, G. & Naraghi, M. Odorant-induced oscillations in the mushroom bodies of the locust. *J. Neurosci.* **14**, 2993–3004 (1994).
32. Pouille, F. & Scanziani, M. Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* **293**, 1159–1163 (2001).
33. Franks, K. M. & Isaacson, J. S. Strong single-fiber sensory inputs to olfactory cortex: implications for olfactory coding. *Neuron* **49**, 357–363 (2006).
34. Luna, V. M. & Schoppa, N. E. GABAergic circuits control input-spike coupling in the piriform cortex. *J. Neurosci.* **28**, 8851–8859 (2008).
35. Ketchum, K. L. & Haberly, L. B. Synaptic events that generate fast oscillations in piriform cortex. *J. Neurosci.* **13**, 3980–3985 (1993).
36. Yaksi, E. & Friedrich, R. W. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved  $\text{Ca}^{2+}$  imaging. *Nature Methods* **3**, 377–383 (2006).
37. Friedrich, R. W. & Korsching, S. I. Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron* **18**, 737–752 (1997).
38. Niessing, J. & Friedrich, R. W. Olfactory pattern classification by discrete neuronal network states. *Nature* **465**, 47–52 (2010).
39. Stokes, C. C. & Isaacson, J. S. From dendrite to soma: dynamic routing of inhibition by complementary interneuron microcircuits in olfactory cortex. *Neuron* **67**, 452–465 (2010).
40. Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: “sparse” coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
41. Stopfer, M., Jayaraman, V. & Laurent, G. Intensity versus identity coding in an olfactory system. *Neuron* **39**, 991–1004 (2003).
42. Jortner, R. A., Farivar, S. S. & Laurent, G. A simple connectivity scheme for sparse coding in an olfactory system. *J. Neurosci.* **27**, 1659–1669 (2007).
43. Spence, R., Gerlach, G., Lawrence, C. & Smith, C. The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol. Rev. Camb. Philos. Soc.* **83**, 13–34 (2008).
44. Engeszer, R. E., Patterson, L. B., Rao, A. A. & Parichy, D. M. Zebrafish in the wild: a review of natural history and new notes from the field. *Zebrafish* **4**, 21–40 (2007).
45. Sato, Y., Miyasaka, N. & Yoshihara, Y. Mutually exclusive glomerular innervation by two distinct types of olfactory sensory neurons revealed in transgenic zebrafish. *J. Neurosci.* **25**, 4889–4897 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Novartis Research Foundation, the Max-Planck-Society, the Swiss National Fonds (SNF), the Deutsche Forschungsgemeinschaft (DFG), the Human Frontier Science Program (HFSP), and the Whitaker Foundation (J.S.). We are grateful to S.-i. Higashijima for vglut2a-GFP transgenic fish and thank T. Frank, A. Lüthi, I. Namekawa and T. Oertner for comments on the manuscript.

**Author Contributions** F.B. performed electrophysiological experiments, analysed data and wrote part of the manuscript. P.Z. generated transgenic fish, participated in the construction of the DMD device, performed optogenetic experiments, recorded LFPs and analysed data. J.S. constructed the DMD device and performed optogenetic experiments. Y.-P.Z.S. performed electrophysiological experiments, participated in the construction of the DMD device, performed calcium imaging experiments, recorded LFPs and analysed data. E.Y. participated in calcium imaging experiments. K.D. contributed channelrhodopsin-2 constructs. R.W.F. conceived the study, designed equipment, analysed data, performed modelling and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.W.F. (Rainer.Friedrich@fmi.ch).



## METHODS

**Animals, experimental preparation and odour delivery.** Experiments were performed in adult zebrafish (*Danio rerio*) that were raised and kept under standard laboratory conditions (26.5 °C; 13/11-h light/dark cycle). All experiments were performed in accordance with official guidelines and approved by the Veterinary Department of the Canton Basel-Stadt (Switzerland). For optical stimulation of sensory inputs to the olfactory bulb, transgenic fish lines were used that express Chr2 (refs 28, 29) selectively in sensory afferents to the olfactory bulb. Five out of thirty-four transgenic fish expressed Chr2 under a 2-kb fragment of the promoter of the zebrafish olfactory marker protein gene<sup>45</sup>. This line expresses Chr2 mainly in sensory afferents projecting to the medial and dorsal olfactory bulb, as described for other transgenic lines generated with this promoter<sup>45</sup>. Fish used in the remaining 29 experiments were generated by expressing Chr2 under the control of a HuC promoter fragment<sup>46</sup> using the Tet system<sup>47</sup>, which results in diverse expression patterns in different founder lines<sup>30</sup>. For experiments in this study, one line was selected that expressed Chr2 in the olfactory bulb selectively in sensory afferents without any obvious regional bias (HuCi:TA/Ptet:Chr2YFP Line 02)<sup>30</sup>. Experiments not including Chr2 stimulation were performed in wild-type fish or fish<sup>48</sup> expressing green fluorescent protein under the control of a promoter of the vesicular glutamate transporter 2a.

Fish were cold-anaesthetized, decapitated and the ventral forebrain was exposed by removal of the jaws and palate. The preparation was transferred into a custom chamber, continuously superfused with teleost artificial cerebrospinal fluid (ACSF)<sup>49</sup>, and warmed up to room temperature. Odours were introduced into a constant stream directed at one naris using a computer-controlled HPLC injection valve. Odour stimuli took approximately 600 ms to reach a plateau concentration and lasted approximately 2.5 s in total. Odour applications were separated in time by at least 40 s, usually longer, to minimize adaptation. Each odour was applied, on average,  $4.1 \pm 1.8$  (mean  $\pm$  s.d.) times and results were averaged, except for the analysis of membrane potential fluctuations and their relationship to LFP oscillations in individual trials, and for power spectral analysis.

**Electrophysiological recordings, calcium imaging and tract stimulation.** Whole-cell patch-clamp recordings were performed using borosilicate pipettes filled with intracellular solution containing (in mM): 130 K-glucuronate, 10 Na-glucuronate, 10 Na-phosphocreatine, 4 NaCl, 4 Mg-ATP, 0.3 Na-GTP, 10 HEPES (pH 7.25;  $\sim$ 300 mosM). For voltage clamp recordings, a Cs-based internal solution was used that contained (in mM): 135 Cs-methanesulphonate, 3 Na-ascorbate, 10 Na-phosphocreatine, 4 MgCl<sub>2</sub>, 4 Na<sub>2</sub>-ATP, 0.4 Na-GTP, 10 HEPES (pH 7.2; 300 mosM). In most experiments, pipette solutions also contained the fluorescent dye Alexa594 to visualize neuronal morphology under two-photon optics. Pipettes had resistances of 10–16 M $\Omega$  and were targeted to somata using contrast-enhanced optics that detected the transmitted infrared laser beam in a two-photon microscope. Signals were recorded at 10 kHz using a Multiclamp 700B amplifier (Molecular Devices) and Scanimage/Ephys software<sup>50,51</sup>. Recording stability was assessed regularly by test pulses. All recorded neurons responded to current injection with action potentials. LFPs were recorded using glass pipettes filled with ACSF and a Multiclamp 700B amplifier.

To stimulate olfactory tracts, glass pipettes filled with ACSF were placed at individual fibre bundles. Because olfactory tracts in zebrafish consist of many dispersed fibre bundles, only a small subset of axons was stimulated by this approach. Stimuli consisted of 10 pulses of 0.5 ms duration at 20 Hz. Intensity (usually 20–30 V) was adjusted to evoke a response of intermediate magnitude.

Multiphoton calcium imaging of odour responses was performed as described<sup>26,36</sup>. Data were acquired at 128 ms per frame. Relative firing rate changes were reconstructed from calcium signals by low-pass filtering and subsequent temporal deconvolution with an exponentially decaying kernel as described<sup>36</sup>. The decay time constant of the kernel ( $\tau_{\text{decay}} = 3$  s) was determined by simultaneous calcium imaging and loose-patch recordings of spikes<sup>36</sup>. No additional filtering before or after deconvolution was performed.

**Optical stimulation.** To generate spatio-temporal patterns of light, the beam of a blue laser (500 mW) was expanded to match the minor diameter (10.5 mm) of a digital micromirror device (DMD) equipped with an accelerator card that allowed for switching each mirror between two angles with a temporal resolution up to 0.125 ms (DMD Discovery 1100 with ALP-I; Texas Instruments). The device was coupled into a two-photon laser scanning microscope using a custom lens system and a dichroic mirror at 45° between the scanners and scan lens. The DMD device was mounted so that the laser light reflected from a given mirror either entered the microscope optics ('on' angle) or was deflected into a beam dump ('off' angle). Light patterns entering the microscope were projected into the image plane through the scan lens, the tube lens, and the objective (20 $\times$ , NA 1.0, Zeiss). Each mirror of the DMD corresponded to a  $0.67 \times 0.67 \mu\text{m}^2$  pixel in the image plane. Light intensity below the objective was  $63 \mu\text{W mm}^{-2}$ . To verify proper resolution, light patterns were projected onto a thin film of fluorescein

solution and fluorescence images were collected using a CCD camera (Supplementary Fig. 2).

Optical stimulation of mitral cells was performed by projecting spatio-temporal patterns of blue light onto the olfactory bulb in fish expressing Chr2 in sensory afferents. In each stimulus, a subset of 'active' squares was switched on and off in a temporal sequence determined by event trains (see below), whereas 'inactive' squares remained off. Two sets of spatial stimulus patterns were generated. Stimulus set 1 consisted of  $50 \times 50 \mu\text{m}^2$  squares in the image plane ( $75 \times 75$  mirrors). Mirrors within each square were switched on and off together, while different squares were controlled by different event trains. Stimulus set 2 consisted of  $20 \times 20 \mu\text{m}^2$  squares ( $30 \times 30$  mirrors), and each mirror was controlled by an individual event train. At each event in a train, the corresponding square (set 1) or mirror (set 2) was switched from the 'off' to the 'on' position for 5 ms to deliver a pulse of blue light to the corresponding pixels in the image plane. The average event rate for each square or mirror was zero before and after the optical stimulus and 20 Hz during stimulation. In set 2, random latencies between zero and 300 ms were assigned to each square to mimic latency differences between responses of different glomeruli. The synchrony between event trains controlling different squares or mirrors was defined by the synchronization index  $S$  (see below). Spatial patterns were created with active squares only over the lateral olfactory bulb, over the medial olfactory bulb, or over the whole olfactory bulb (Supplementary Fig. 2). On average, each optical stimulus was repeated  $3.4 \pm 1.2$  times (mean  $\pm$  s.d.; mitral cells,  $3.4 \pm 1.1$  repetitions; Dp neurons,  $3.5 \pm 1.2$  repetitions) and results were averaged (except for power spectral analysis).

Responses to stimulus set 2 ( $n = 68$  responses from 21 mitral cells and  $n = 55$  responses from 18 Dp neurons at each  $S$ ) increased more slowly than responses to stimulus set 1 ( $n = 93$  responses from 28 mitral cells and  $n = 38$  responses from 16 Dp neurons) because squares were activated successively during the initial 300 ms. However, the time difference between responses of mitral cell and Dp neurons was similar in response to both stimulus sets: the mean peak time of strong depolarizations ( $>10$  s.d.) in response to stimulus set 1 was  $108 \pm 48$  ms for mitral cells and  $263 \pm 110$  ms for Dp neurons. The mean peak time of strong depolarizations ( $>10$  s.d.) in response to stimulus set 2 was  $261 \pm 125$  ms for mitral cells and  $438 \pm 178$  ms for Dp neurons. The mean peak time of strong firing rate changes ( $>10$  s.d.) in response to stimulus set 1 was  $132 \pm 179$  ms for mitral cells and  $283 \pm 213$  ms for Dp neurons. The mean peak time of strong firing rate changes ( $>10$  s.d.) in response to stimulus set 2 was  $267 \pm 228$  ms for mitral cells and  $407 \pm 206$  ms for Dp neurons. Otherwise, responses to both stimulus sets were similar. Data obtained with the two stimulus sets were therefore pooled.

The stability of mitral cell firing rates to changes in synchrony may be due to the fact that mitral cell membrane potentials are near spike threshold already in the absence of odour stimulation so that synchrony affects primarily the timing of action potentials. Moreover, stronger mitral cell activation is counterbalanced by increased feedback inhibition in the olfactory bulb<sup>52</sup>, and inputs of different strengths might be equalized by a nonlinear transformation at the sensory neuron-to-mitral cell synapse, as found in *Drosophila*<sup>53</sup>.

**Generation of event trains and modelling.** To generate trains of events (spike times or DMD mirror activations) with varying degrees of synchrony, individual events were drawn from probability distribution  $P(t)$  given by

$$P(t) = X(t) / \int_0^{t_{\text{max}}} X(t) dt \text{ with}$$

$$X(t) = (S \times \sin(20 \times 2\pi \times t) + 1) \quad | S \leq 1$$

$$X(t) = (\sin(20 \times 2\pi \times t) + 1)^S \quad | S \leq 1$$

where  $t \leq t_{\text{max}}$  is time in seconds and  $S$  is the synchronization index. To allow for a refractory period, events closer than 5 ms to an event kept previously were discarded. This procedure was repeated until the mean number of events per unit time was equal to the desired event rate. For  $S = 0$ , the instantaneous probability of an event was defined to be constant. This procedure parameterized synchrony in simulated spike trains and in optical stimulus patterns.

Event trains were transformed into excitatory postsynaptic currents by convolving the delta functions representing event times with an alpha function representing a unitary excitatory postsynaptic current (EPSC; rise time, 0.8 ms; decay time, 5 ms).

To simulate the integration of synaptic currents in a simple model, a conductance-based, non-spiking point neuron model without voltage-gated conductances was used. The input resistance  $R_{\text{in}}$  was  $2 G\Omega$  and the time constant was  $\tau_{\text{m}} = 40$  ms, similar to values measured in Dp neurons. The neuron received  $n = 10$  inputs, each firing at a mean rate of 10 Hz. Each input spike evoked excitatory and inhibitory synaptic currents with a time course (250 ms length) given by the mean current measured at these potentials in Dp neurons in response to electrical stimulation of the olfactory tract (Supplementary Fig. 13). Since time courses of

synaptic currents could be measured only over 50 ms between successive stimulus pulses, the remaining 200 ms were extrapolated by exponential functions. Reversal potentials of excitatory and inhibitory synaptic currents were  $V_{\text{rev,exc}} = 0$  mV and  $V_{\text{rev,inh}} = -70$  mV, respectively. The evolution of the membrane potential  $V_m$  was then given by

$$CdV_m/dt = -g_0(V(t) - V_{\text{rev},0}) - g_{\text{exc}}(t)(V(t) - V_{\text{rev,exc}}) - g_{\text{inh}}(t)(V(t) - V_{\text{rev,inh}})$$

where  $C = \tau_m/R_{\text{in}}$  is membrane capacitance,  $g_0 = 1/R_{\text{in}}$  is the input conductance,  $V_{\text{rev},0} = -70$  mV is the resting potential,  $g_{\text{exc}}(t)$  is the excitatory synaptic conductance and  $g_{\text{inh}}(t)$  is the inhibitory synaptic conductance.

Simulations were performed for input spike patterns with three different synchronies  $S$ . For each synchrony, we also shifted the onset of inhibitory currents backward and forward in time ( $-5$  ms and  $+20$  ms). Moreover, we varied the mean firing rate of each input between 8 Hz and 12 Hz, and the number of inputs  $N$  between 8 and 12 (without changing the amplitude of synaptic currents). For each condition, results of 128 simulations, each 2.5 s long, were averaged.

**Data analysis.** Power spectral analysis was performed on individual trials. Power spectra were then first averaged over trials with the same stimulus, and subsequently averaged over stimulus–neuron pairs. LFPs in the olfactory bulb were recorded in the deep layers, beyond the point of reversal<sup>24</sup>. Oscillations were analysed by band-pass filtering LFP recordings between 5 and 40 Hz using a non-causal filter designed in Matlab (The MathWorks). Spike phases were determined by the time of each spike relative to the neighbouring peaks in the band-pass filtered LFP. Spike phase distributions were determined for each neuron–stimulus pair from repeated trials and normalized by the number of trials and spikes. Experiments in which less than 16 spikes were collected in total were excluded. Distributions were then averaged over neuron–stimulus pairs.

Trains of action potentials were transformed into continuous firing rates by convolution with a Gaussian kernel (s.d., 5 ms). Mean changes in membrane potential and mean firing rates in response to optical stimulation were determined within a 1-s time window following stimulus onset. For analyses of membrane potential dynamics, action potentials were removed by a sliding median filter with width 14 ms. To determine peak times of membrane potential responses, the time course of the median-filtered membrane potential trace was smoothed by a non-causal four-pole low-pass Butterworth filter with cutoff frequency 50 Hz.

The resting potential was determined as the mean over the 500 lowest membrane potential values (1-ms bins) during a 1.5–3 s pre-odour period. Action

potential thresholds were determined as the inflection point of action potentials in each trial. To determine the amplitude of membrane potential fluctuations in the 5–40 Hz band and of the slow depolarization during odour stimulation, membrane potential traces were first low-pass filtered by a non-causal four-pole Butterworth filter with cutoff frequency 5 Hz. The amplitude of the slow depolarization was quantified as the difference between the resting potential and the peak of the low-pass filtered membrane potential trace. Fluctuations were isolated by subtracting the low-pass filtered trace from the raw trace and subsequently band-pass filtered with the same filter as used for the LFP. The amplitudes of fluctuations were then quantified as the amplitudes of the positive peaks. Maximal amplitudes of fluctuations in a given trial were determined as the mean over the 10% largest fluctuations. For the separate analysis of subthreshold and suprathreshold responses, subthreshold responses were defined as neuron–stimulus pairs that never responded with action potentials to a given odour, whereas suprathreshold responses were defined as responses that included, on average, at least one spike.

Significance tests were performed using a non-parametric Wilcoxon rank-sum test unless noted otherwise.

46. Higashijima, S., Masino, M. A., Mandel, G. & Fetcho, J. R. Imaging neuronal activity during zebrafish behavior with a genetically encoded calcium indicator. *J. Neurophysiol.* **90**, 3986–3997 (2003).
47. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551 (1992).
48. Miyasaka, N. *et al.* From the olfactory bulb to higher brain centers: genetic visualization of secondary olfactory pathways in zebrafish. *J. Neurosci.* **29**, 4756–4767 (2009).
49. Mathieson, W. B. & Maler, L. Morphological and electrophysiological properties of a novel in vitro preparation: the electrosensory lateral line lobe brain slice. *J. Comp. Physiol. A* **163**, 489–506 (1988).
50. Polgruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
51. Suter, B. A. *et al.* Ephus: multipurpose data acquisition software for neuroscience experiments. *Front. Neural Circuits* **4**, 100 (2010).
52. Tabor, R. & Friedrich, R. W. Pharmacological analysis of ionotropic glutamate receptor function in neuronal circuits of the zebrafish olfactory bulb. *PLoS ONE* **3**, e1416 (2008).
53. Bhandawat, V., Olsen, S. R., Gouwens, N. W., Schlieff, M. L. & Wilson, R. I. Sensory processing in the *Drosophila* antennal lobe increases reliability and separability of ensemble odor representations. *Nature Neurosci.* **10**, 1474–1482 (2007).



# The alignment of molecular cloud magnetic fields with the spiral arms in M33

Hua-bai Li<sup>1</sup> & Thomas Henning<sup>1</sup>

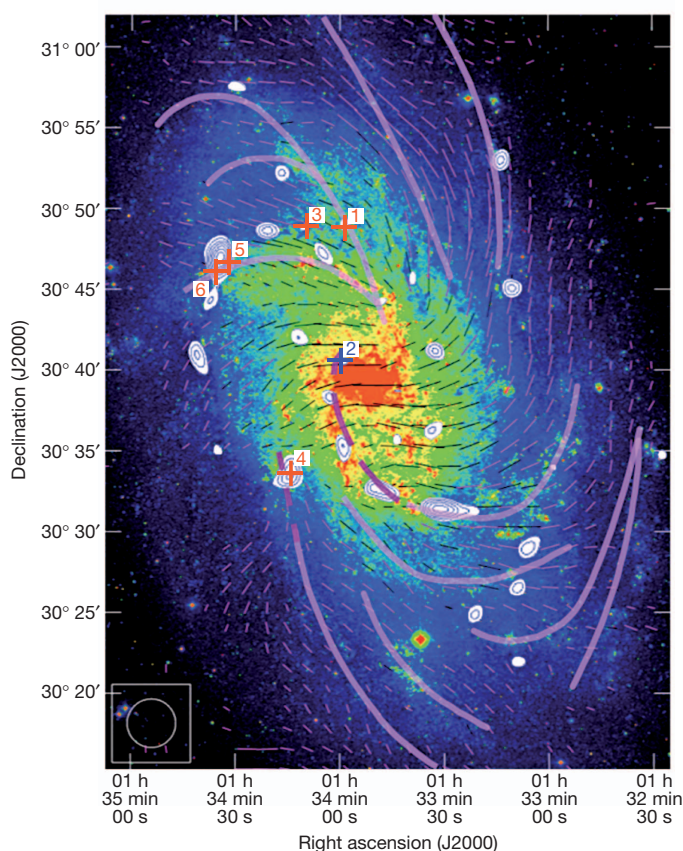
**The formation of molecular clouds, which serve as stellar nurseries in galaxies, is poorly understood. A class of cloud formation models suggests that a large-scale galactic magnetic field is irrelevant at the scale of individual clouds, because the turbulence and rotation of a cloud may randomize the orientation of its magnetic field<sup>1,2</sup>. Alternatively, galactic fields could be strong enough to impose their direction upon individual clouds<sup>3,4</sup>, thereby regulating cloud accumulation and fragmentation<sup>5</sup>, and affecting the rate and efficiency of star formation<sup>6</sup>. Our location in the disk of the Galaxy makes an assessment of the situation difficult. Here we report observations of the magnetic field orientation of six giant molecular cloud complexes in the nearby, almost face-on, galaxy M33. The fields are aligned with the spiral arms, suggesting that the large-scale field in M33 anchors the clouds.**

At a distance of about 900 kiloparsecs (kpc) from us<sup>7</sup>, M33 is our nearest face-on galaxy with pronounced optical spiral arms. To resolve a typical giant molecular cloud (GMC) with a size of tens to a hundred parsecs, we used the Submillimeter Array<sup>8</sup>, which offers a linear spatial resolution of about 15 pc at 230 GHz (the frequency of the CO J = 2–1 transition) at the distance of M33 using the array's most compact configuration. To observe the strongest CO line emission, we picked the six most massive GMCs from M33 (ref. 9). It is clear to which spiral arms the GMCs are related, except for GMC3, which is situated between two optical arms (Fig. 1).

We determined the orientations of the GMC magnetic fields ( $B$  fields) from the polarization of CO emission lines, which should be either perpendicular or parallel to the local  $B$ -field direction projected on the sky (the Goldreich–Kylafis effect<sup>10</sup>). Although there are other  $B$ -field tracers<sup>11</sup> that do not have this 90° ambiguity, CO is much more abundant and allows current radio telescopes to perform extragalactic cloud observations. Despite the 90° ambiguity, such a  $B$ -field observation is still valuable<sup>12</sup>. An intrinsically random field distribution, as occurs when the turbulence is super-Alfvénic<sup>13</sup> (that is, when turbulent energy dominates  $B$ -field energy), will still be random with this ambiguity. On the other hand, an intrinsically single-peaked Gaussian-like field distribution, in the presence of sub-Alfvénic turbulence<sup>13</sup> (that is, when  $B$ -field energy dominates turbulent energy), will either remain single-peaked, or split into two peaks approximately 90° apart ('double peaks'). From the total distribution of the offsets between the CO polarization of the M33 GMCs and the local arm directions (Fig. 2), the trend of double peaks is clearly visible (Fig. 3). The distribution can be fitted by a double-Gaussian function with peaks at  $-1.9^\circ \pm 4.7^\circ$  and  $91.1^\circ \pm 3.7^\circ$  and a standard deviation of  $20.7^\circ \pm 2.6^\circ$ . This result is barely affected if the inter-arm GMC3 is excluded.

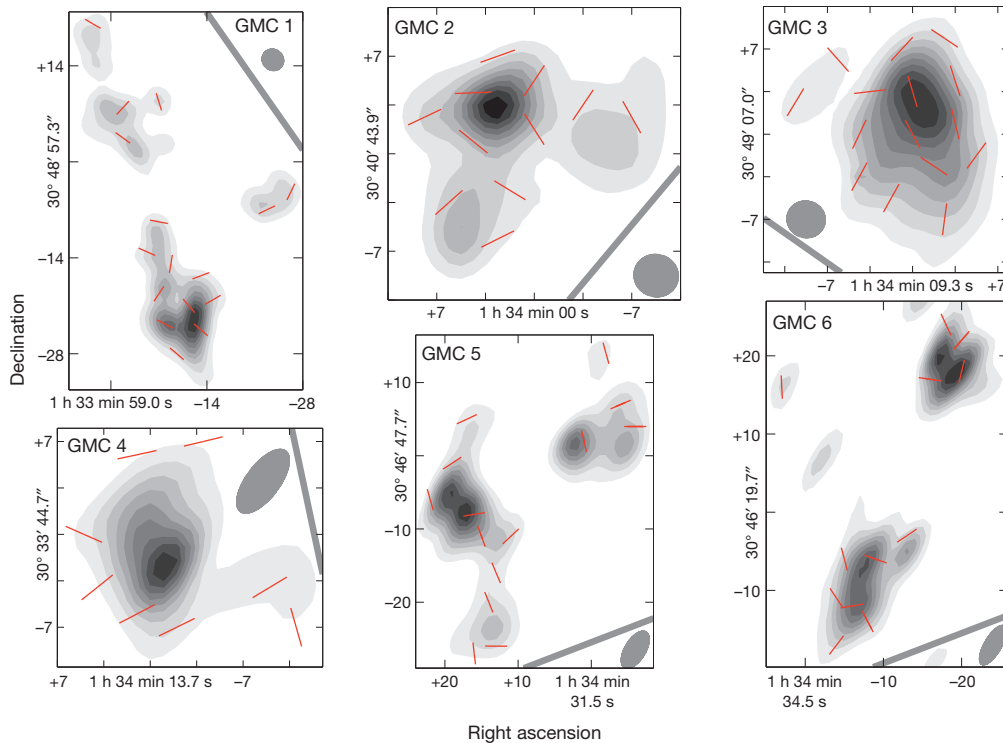
The angle dispersion ( $\Sigma$ ) of the CO polarization and its offset ( $\Delta$ ) from the arm directions are determined by the dispersion of  $B$ -field orientations ( $\sigma$ ), the dispersion of the offsets between cloud mean fields and spiral arms ( $\delta$ ), the observational errors ( $\varepsilon < 10^\circ$ ), and the Goldreich–Kylafis effect. By performing Monte Carlo simulations (see the Supplementary Information), we can estimate the likelihood of observing  $\Sigma$  and  $\Delta$  within certain ranges, and determine which

combinations of  $\sigma$  and  $\delta$  are able to produce the observed confidence level (Fig. 4). Random fields or field-arm offsets are very unlikely. Only when  $\sigma = 17^\circ$ – $22^\circ$  with  $\delta < 8^\circ$  do the simulations give a confidence level similar to that observed. This indicates that the mean field directions are well-defined and highly correlated with the spiral arms, which



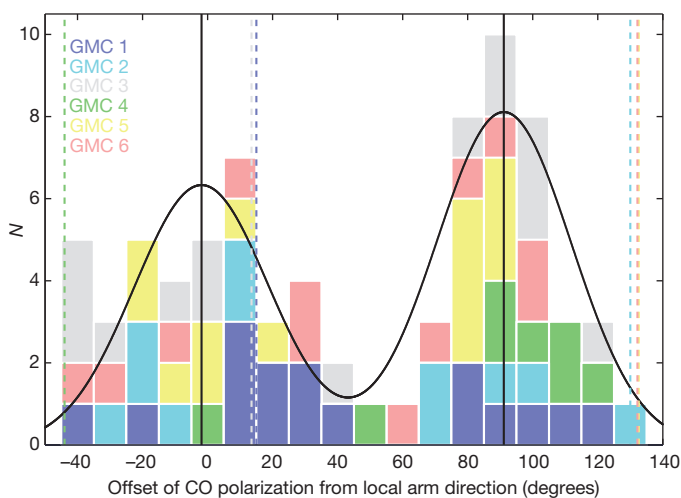
**Figure 1 | The optical spiral arms and the locations of the six most massive GMCs in M33.** The background is an optical image of M33 (ref. 21) and the vectors show the 3.6-cm synchrotron polarization<sup>21</sup>, with the telescope beam size shown in the lower left corner. The vector length is proportional to the intensity of polarized emission. A sketch of the optical arms<sup>27</sup> is shown with thick purple solid lines. The contours show the structures between 36'' and 48'' derived from scale decomposition<sup>28</sup> of the 500-μm Herschel data<sup>29</sup> (the lowest contour level has 30% of the peak intensity, and the following levels increase linearly with steps of 10% peak intensity). The GMC locations (plus symbols) are numbered 1 to 6. The optical arms related to GMCs 1, 5 and 6 are clear. GMC 3 is between two arms. GMCs 2 and 4 are on the extensions (purple dashed lines) from two different solid lines. GMC 4 has a short straight extension. For GMC 2, we adopt the southern arm defined by ref. 30 as the extension, which traces the 500-μm clumps well. The arms are traced slightly differently in the literature, and this observational uncertainty will contribute to  $\delta$ , the dispersion of polarization-arm offset (Supplementary Information).

<sup>1</sup>Max-Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany.



**Figure 2 | CO (2-1) maps and polarization vectors.** The contours are 90%, 80% ... 10% of the peak intensity of each cloud (from dark to light grey). The red vectors show polarization detections, for which the ratio of the polarization level to its uncertainty is greater than three and the error in direction is less than  $10^\circ$ . The thick grey vectors show the tangents of the local optical arms<sup>27,30</sup> for

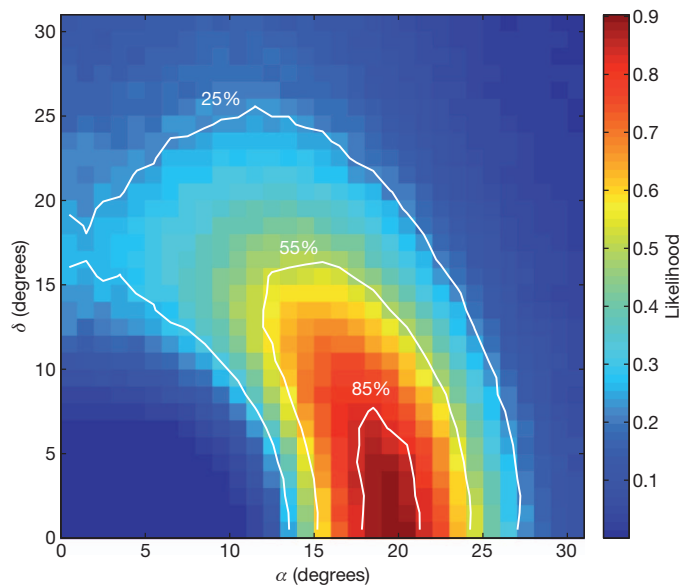
is consistent with the scenario in which the  $B$  fields in the GMCs are compressed within the spiral arms, and the fields can exert tension forces (because of  $\delta$ ; see Supplementary Fig. 1) strong enough to resist cloud rotation ('magnetic braking'<sup>4,9</sup>). If mass has been accreted to form a cloud from the accumulation length scale (hundreds of parsecs) in a shearing galactic disk, cloud rotation is inevitable owing to conservation of angular momentum (Supplementary Fig. 1 and figure 6 of



**Figure 3 | Distribution of the CO polarization-arm offsets.** The offsets are from the difference between the red and grey vectors in Fig. 2.  $N$  is the number of offsets falling in each bin. Contributions from different GMCs are indicated by colour. The distribution can be fitted by a double-Gaussian function with a standard deviation of  $20.7^\circ \pm 2.6^\circ$  and peaks at  $-1.9^\circ \pm 4.7^\circ$  and  $91.1^\circ \pm 3.7^\circ$  (peaks shown by vertical black lines). The directions of synchrotron polarization from the regions near the GMCs (within one beam size, as shown in Fig. 1) are also shown as dashed lines, with the same colour code as the GMCs.

GMC 1, 2, 4, 5 and 6. For the inter-arm GMC 3, the grey vector shows the mean of the two tangents of the nearby arms at the positions closest to GMC 3. The ellipses indicate the Submillimeter Array synthetic beams; all detections are spatially independent. The coordinate (right ascension, declination) offsets are in arcseconds.

ref. 2), unless the momentum is consumed by another mechanism such as magnetic tension. The fact that the GMCs in M33 show significantly smaller angular momenta than predicted by the Toomre instability criterion<sup>9</sup> supports our observations. Figure 3 is also consistent with



**Figure 4 | Likelihood of obtaining simulated angle dispersions  $\sigma$  and offsets  $\delta$  of CO polarization within the observed 90% confidence intervals.** The 90% confidence intervals from the data in Fig. 3 are  $15.1^\circ < \sigma < 26.4^\circ$  and  $|\delta| < 10.7^\circ$ . The likelihood is estimated by Monte Carlo simulations (Supplementary Information) with various combinations of  $\sigma$  ( $B$ -field dispersion) and  $\delta$  (dispersion of the offsets between mean fields and arms). Only  $\sigma \approx 20^\circ$  with  $\delta < 8^\circ$  can give a similar confidence level within the intervals.



the cloud  $B$  fields being perpendicular to the arms; however, there is neither a known mechanism nor a numerical simulation supporting such a field configuration.

The CO polarization is not necessarily aligned with the local synchrotron polarization (Fig. 3), which traces the  $B$  fields in regions near the GMCs that are warmer and more diffuse<sup>14</sup> than the clouds themselves. From the viewpoint of the dynamo theory, it is possible for cloud fields to decouple from the large-scale galactic fields owing to the small volume-filling factor of molecular clouds<sup>15</sup>. This is consistent with the decoupling of synchrotron/CO polarization, but fails to explain the correlation between cloud  $B$  fields and spiral arms. To our knowledge, no numerical simulations of small-scale dynamos within molecular clouds have been able to produce ordered field directions coherent with spiral arms.

However, density wave compression, the cause of spiral arms, can explain the observation as follows. In the classical picture of interstellar media<sup>16</sup>, cold and warmer phases are in pressure equilibrium. While orbiting through the gravitational well of the spiral arms, the cold phase experiences a much stronger shock compression than does the warmer phase<sup>17</sup>. Therefore, it is unsurprising that the orientation of the  $B$  fields in the warmer media (traced by synchrotron radiation) can decouple from the morphology of the more compressed fields in cold media (traced by CO). Because the gas and  $B$  fields in the cold media are compressed together, they should follow the same spiral morphology as long as the subsequent cloud formation activity does not disturb the field orientations significantly.

The non-random  $B$  fields imply that the cloud turbulence is sub-Alfvénic. For this kind of turbulence, we can follow Chandrasekhar and Fermi<sup>18</sup>, assuming that the dispersion in  $B$ -field direction is coupled with the lateral velocity of gas turbulence, to estimate the  $B$ -field strength:  $B_{\text{pos}} = 9.3 \sqrt{n(\text{H}_2)(v/\sigma)}$ , where  $B_{\text{pos}}$ ,  $n(\text{H}_2)$  and  $v$  are, respectively, the plane-of-sky component of the  $B$  field (in  $\mu\text{G}$ ),  $\text{H}_2$  density (in  $\text{cm}^{-3}$ ), and full-width at half-maximum (FWHM) spectral linewidth (in  $\text{km s}^{-1}$ ). This equation is modified from the original one in ref. 18 after calibration with numerical simulations<sup>19</sup>. With  $\sigma = 17^\circ\text{--}22^\circ$ ,  $n(\text{H}_2) = 10^3\text{--}10^4 \text{ cm}^{-3}$  (ref. 20), and  $v \approx 10 \text{ km s}^{-1}$  (ref. 9), the  $B_{\text{pos}}$  value is 0.1–0.5 mG. This value is comparable to that of  $B$  fields in the molecular clouds of the Milky Way, and is about 100 times greater than the  $B$ -field strength estimated from synchrotron observations<sup>21</sup>.

Evidence of sub-Alfvénic turbulence is also observed within the clouds of the Milky Way. The Galactic  $B$  field also anchors molecular clouds<sup>11,22</sup> and aligns the velocity anisotropy of turbulence<sup>5,23</sup>. From our edge-on view of the disk, the Milky Way fields, however, have rich structures at the hundred-parsec scale (the scale of the GMCs), instead of being aligned with the disk<sup>11</sup>. If galactic dynamics are similar for the Milky Way and M33, the simplest explanation is that  $B$  fields of spiral arms can have much more structure perpendicular to the disk (as observed from an edge-on view) than within the disk plane (as observed from a bird's-eye view). Several mechanisms can help generate this anisotropic  $B$ -field structure. First, density wave compression occurs mainly in a direction parallel to the disk plane, and second, the Parker instability<sup>24</sup> and stellar feedback (for example, bipolar giant H II bubbles<sup>25</sup> and galactic fountains<sup>26</sup>) concentrate along gaseous arms, tending to deviate the  $B$  fields towards directions perpendicular to the disk. A bird's-eye view of cloud  $B$  fields is currently much more difficult to acquire compared to an edge-on view, but could offer important new insights into GMC/galaxy dynamics. Next-generation array telescopes (for example, the Atacama Large Millimeter/submillimeter Array) will be better able to survey molecular clouds efficiently from different views.

Received 16 March; accepted 6 September 2011.

Published online 16 November 2011.

1. Dobbs, C. GMC formation by agglomeration and self gravity. *Mon. Not. R. Astron. Soc.* **391**, 844–858 (2008).

2. Hartmann, L., Ballesteros-Paredes, J. & Bergin, E. Rapid formation of molecular clouds and stars in the solar neighborhood. *Astrophys. J.* **562**, 852–868 (2001).
3. Passot, T., Vazquez-Semadeni, E. & Pouquet, A. A turbulent model for the interstellar medium. II. Magnetic fields and rotation. *Astrophys. J.* **455**, 536–555 (1995).
4. Shetty, R. & Ostriker, E. Global modeling of spur formation in spiral galaxies. *Astrophys. J.* **647**, 997–1017 (2006).
5. Li, H.-b. et al. Evidence for dynamically important magnetic fields in molecular clouds. *Mon. Not. R. Astron. Soc.* **411**, 2067–2075 (2011).
6. Price, D. & Bate, M. The effect of magnetic fields on star cluster formation. *Mon. Not. R. Astron. Soc.* **385**, 1820–1834 (2008).
7. Kim, M., Kim, E., Lee, M., Sarajedini, A. & Geisler, D. Determination of the distance to M33 based on the tip of the red giant branch and the red clump. *Astron. J.* **123**, 244–254 (2002).
8. Ho, P., Moran, J. M. & Lo, K. Y. The submillimeter array. *Astrophys. J.* **616**, L1–L6 (2004).
9. Rosolowsky, E., Engargiola, G., Plambeck, R. & Blitz, L. Giant molecular clouds in M33. II. High-resolution observations. *Astrophys. J.* **599**, 258–274 (2003).
10. Goldreich, P. & Kylafis, N. On mapping the magnetic field direction in molecular clouds by polarization measurements. *Astrophys. J.* **243**, L75–L78 (1981).
11. Li, H.-b., Dowell, C., Goodman, A., Hildebrand, R. & Novak, G. Anchoring magnetic field in turbulent molecular clouds. *Astrophys. J.* **704**, 891–897 (2009).
12. Greaves, J., Holland, W., Friberg, P., Dent, W. & Polarized, C. O. Emission from molecular clouds. *Astrophys. J.* **512**, L139–L142 (1999).
13. Falceta-Gonçalves, D., Lazarian, A. & Kowal, G. Studies of regular and random magnetic fields in the ISM: statistics of polarization vectors and the Chandrasekhar-Fermi technique. *Astrophys. J.* **679**, 537–551 (2008).
14. Frick, P., Beck, R., Berkhuijsen, E. & Patrickeyev, I. Scaling and correlation analysis of galactic images. *Mon. Not. R. Astron. Soc.* **327**, 1145–1157 (2001).
15. Shukurov, A. in *Dynamic Interstellar Medium: Recent Numerical Simulations. Plasma Turbulence and Energetic Particles in Astrophysics* (eds Ostrowski, M. & Schlickeiser, R.) 66–73 (Observatorium Astronomiczne, Uniwersytet Jagielloński, 1999).
16. Field, G., Goldsmith, D. & Habing, H. G. Cosmic-ray heating of the interstellar gas. *Astrophys. J.* **155**, L149–L154 (1969).
17. Kim, C.-G., Kim, W.-T. & Ostriker, E. Galactic spiral shocks with thermal instability. *Astrophys. J.* **681**, 1148–1162 (2008).
18. Chandrasekhar, S. & Fermi, E. Magnetic fields in spiral arms. *Astrophys. J.* **118**, 113–115 (1953).
19. Crutcher, R. In *The Magnetized Plasma in Galaxy Evolution* (eds Chyży, K., Otmianowska-Mazur, K., Soida, M. & Dettmar, R.-J.) 103–110 (2005).
20. Wilson, C., Walker, C. & Thornley, M. The density and temperature of molecular clouds in M33. *Astrophys. J.* **483**, 210–219 (1997).
21. Tabatabaei, F., Krause, M., Fletcher, A. & Beck, R. High-resolution radio continuum survey of M33. III. Magnetic fields. *Astron. Astrophys.* **490**, 1005–1017 (2008).
22. Han, J. & Zhang, J. The Galactic distribution of magnetic fields in molecular clouds and H II regions. *Astron. Astrophys.* **464**, 609–614 (2007).
23. Heyer, M., Gong, H., Ostriker, E. & Brunt, C. Magnetically aligned velocity anisotropy in the Taurus molecular cloud. *Astrophys. J.* **680**, 420–427 (2008).
24. Mouschovias, T., Kunz, M. & Christie, D. Formation of interstellar clouds: Parker instability with phase transitions. *Mon. Not. R. Astron. Soc.* **397**, 14–23 (2009).
25. Li, H.-b. et al. Results of SPARO 2003: mapping magnetic fields in giant molecular clouds. *Astrophys. J.* **648**, 340–354 (2006).
26. Kwak, K., Shelton, R. & Raley, E. The evolution of gas clouds falling in the magnetized Galactic halo: high-velocity clouds (HVCs) originated in the Galactic fountain. *Astrophys. J.* **699**, 1775–1788 (2009).
27. Sandage, A. & Humphreys, R. On the warped optical plane of M33. *Astrophys. J.* **236**, L1–L5 (1980).
28. Men'shchikov, A. et al. Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel. *Astron. Astrophys.* **518**, L103–L107 (2010).
29. Kramer, C. et al. PACS and SPIRE photometer maps of M33: first results of the HERschel M33 Extended Survey (HERM33ES). *Astron. Astrophys.* **518**, L67–L71 (2010).
30. Rogstad, D., Wright, M. & Lockhart, I. Aperture synthesis of neutral hydrogen in the galaxy M33. *Astrophys. J.* **204**, 703–716 (1976).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Rosolowsky, R. Shetty, T. K. Sridharan, M. Houde, S. Paine, H.-H. Wang, A. Karim, S. Ragan, K. Smith, P. Boley and T. Wu for comments. We appreciate the help of D. Marrone, G. Petitpas and R. Rao with the observations. We are grateful for the Herschel maps of M33 offered by C. Kramer. This research is supported by the Max-Planck-Institut für Astronomie and Harvard-Smithsonian Center for Astrophysics. The Submillimeter Array is a joint project between the Smithsonian Astrophysical Observatory and the Academia Sinica Institute of Astronomy and Astrophysics and is funded by the Smithsonian Institution and the Academia Sinica.

**Author Contributions** H.L. designed and executed the experiment. H.L. and T.H. contributed jointly to the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.L. (li@mpia.de).

# Active formation of ‘chaos terrain’ over shallow subsurface water on Europa

B. E. Schmidt<sup>1</sup>, D. D. Blankenship<sup>1</sup>, G. W. Patterson<sup>2</sup> & P. M. Schenk<sup>3</sup>

Europa, the innermost icy satellite of Jupiter, has a tortured young surface<sup>1–4</sup> and sustains a liquid water ocean<sup>1–6</sup> below an ice shell of highly debated thickness<sup>1–5,7–10</sup>. Quasi-circular areas of ice disruption called chaos terrains are unique to Europa, and both their formation and the ice-shell thickness depend on Europa’s thermal state<sup>1–5,7–17</sup>. No model so far has been able to explain why features such as Conamara Chaos stand above surrounding terrain and contain matrix domes<sup>10,18</sup>. Melt-through of a thin (few-kilometre) shell<sup>3,7,8</sup> is thermodynamically improbable and cannot raise the ice<sup>10,18</sup>. The buoyancy of material rising as either plumes of warm, pure ice called diapirs<sup>1,9–15</sup> or convective cells<sup>16,17</sup> in a thick (>10 kilometres) shell is insufficient to produce the observed chaos heights, and no single plume can create matrix domes<sup>10,18</sup>. Here we report an analysis of archival data from Europa, guided by processes observed within Earth’s subglacial volcanoes and ice shelves. The data suggest that chaos terrains form above liquid water lenses perched within the ice shell as shallow as 3 kilometres. Our results suggest that ice–water interactions and freeze-out give rise to the diverse morphologies and topography of chaos terrains. The sunken topography of Thera Macula indicates that Europa is actively resurfacing over a lens comparable in volume to the Great Lakes in North America.

Although the settings are different, terrestrial environments can provide critical context for Europa, particularly where water and ice interact under pressure. Melting of ice occurs at subglacial volcanic craters on Earth, such as Iceland’s Grimsvotn<sup>19</sup> (Supplementary Fig. 4). Ice covers the volcano, and as it becomes active, the ice cap melts from below, causing surface down draw<sup>19</sup>. Water in glacial systems flows perpendicular to the gradient of the fluid potential (Supplementary Information section 2). In pure ice, the surface slope above a water body is roughly 11 times as important as the basal slope in determining flow direction<sup>19,20</sup>, and water collects where the hydraulic gradient approaches zero, at the centre of the depression. Above the activating volcano, the surface slope is steepest at the flanks of the crater, creating a hydraulic seal that prevents the escape of water and drives the formation of a lens-shaped subglacial lake<sup>19,20</sup>. Because Iceland’s ice caps are finite in width, breakout flow along the bed or flotation of the ice cap eventually allows water to escape<sup>19,20</sup>.

Water also influences Antarctic ice shelves. Brines enter terrestrial ice shelves through basal fractures or the front of the shelf (via a porous layer called firn) and percolate through the ice for tens of kilometres over many years<sup>21</sup>. Beyond enhancing its water and impurity content, introduction of brine can weaken the ice by reducing its shear strength<sup>21–23</sup>. Hydrofracture occurs when tidal cracks fill with water, causing force at the crack tip, which can initiate ice shelf collapse<sup>22,23</sup> (Supplementary Information section 2), producing tabular icebergs surrounded by a matrix of brine-rich (‘brash’) ice.

Coupled analysis of the geomorphology of Conamara Chaos and Thera Macula, (Figs 1 and 2, respectively) demonstrates that the two features share a quasi-circular shape and floating blocks, whereas their

topography differs. Conamara Chaos is raised and contains raised matrix ‘domes’. Thera Macula, however, is sunken below the surrounding surface. These observations, informed by the environments on Earth described above, suggest a four-phase ‘lens-collapse model’ for chaos terrain formation (Fig. 3). (1) Ascending thermal plumes of relatively pure ice<sup>13</sup> cross the eutectic point of overlying impure ice, producing surface deflection in response to volume change associated with pressure melting of the ice (Fig. 3a). (2) Resulting hydraulic gradients and driving forces produce a sealed, pressurized melt lens (Fig. 3b). (3) Extension of the sinking brittle ice ‘lid’ over the lens ultimately generates deep fractures from below, allowing brine to both be injected into and percolate through overlying ice, forming a fluidized granular ice matrix and calving ice blocks (Fig. 3c). (4) Refreezing of the melt lens and now brine-rich matrix results in topographic heterogeneity (Fig. 3d).

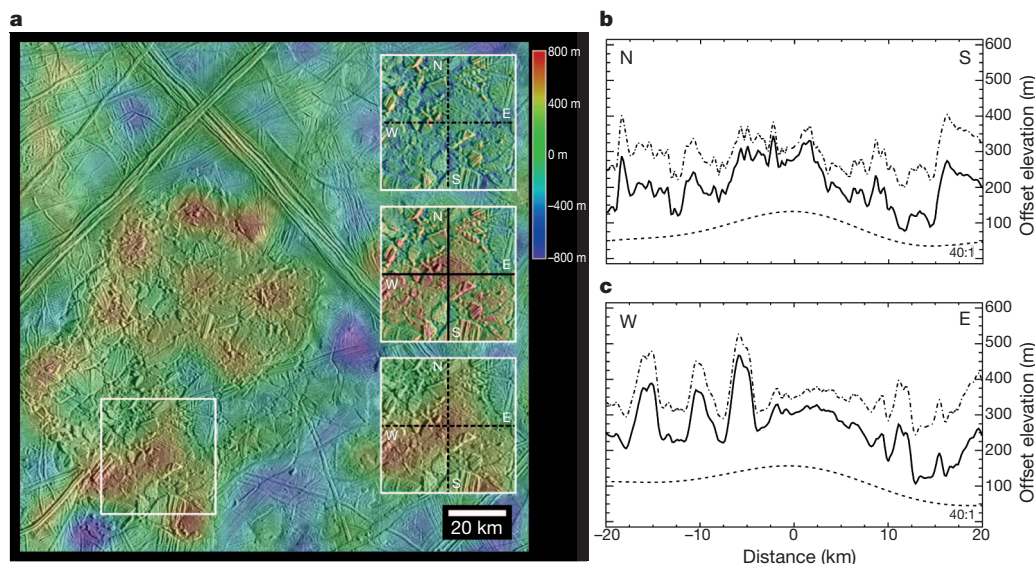
The upper crust of Europa is rich in impurities owing to either exogenic implantation or endogenic injection (see, for example, refs 24, 25). Models suggest that melt will be formed as warm, compositionally buoyant plumes cause the cold impurity-rich ice above them to reach its eutectic pressure-melt point, driving partial melting and ice disruption<sup>1,2,9–15,17</sup>. Interconnected pockets of thermally stable melt water will form above a plume and be over-pressurized by an amount equal to the buoyant force of the plume,  $10^4$ – $10^5$  Pa (ref. 13). It is significant that previous models did not take into account the volume change associated with melting ice<sup>15</sup> and its ramifications<sup>19,20</sup>. Ice melting results in surface draw down, which hydraulically seals the melt in place<sup>19,20</sup> (Fig. 3b; Supplementary Information section 2), rather than melt draining downward as previously speculated<sup>10,13–15,17,26</sup>. That is, water formed from melting above plumes must move perpendicular to local hydraulic gradients, both up the plume head and towards the centre of the depression<sup>24</sup>, and will form a lens similar to subglacial volcanic lakes. Ultimately, the volume of melt and hydrostatic equilibrium with the overlying depression determine the shape of the lens; this shape will be slightly modified by any lithostatic stresses at the edge of the lens. On Europa, the lateral continuity of the ice shell prohibits horizontal escape of the water.

Pressure and fracture will contribute to the response of Europa’s ice shell to subsurface melting. Prevalent pre-existing faults and discontinuities in ice strength should be regions of localized weakness, akin to ice shelf fractures, accumulating much of the extensional strain from the subsidence of the lid. As cracks propagate upward from the melt lens, hydrofracture will break up the ice (Fig. 3c) wherever high-pressure water inflow contributes force at the crack tip<sup>22,23</sup>. Fracture will calve steep-sided blocks and allow water to enter overlying brittle ice.

The morphology of Conamara Chaos requires the transformation of mostly background plains material<sup>3,11,14</sup> into an impurity-rich matrix (Supplementary Figs 1, 6). Background plains material is characterized by high fracture density, and thus may be more susceptible to brine inflow and disruption as observed<sup>11,14</sup>; the shallow subsurface may also

<sup>1</sup>Institute for Geophysics, John A. & Katherine G. Jackson School of Geosciences, The University of Texas at Austin, J. J. Pickle Research Campus, Building 196 (ROC), 10100 Burnet Road (R2200), Austin, Texas 78758-4445, USA. <sup>2</sup>Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA. <sup>3</sup>Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA.



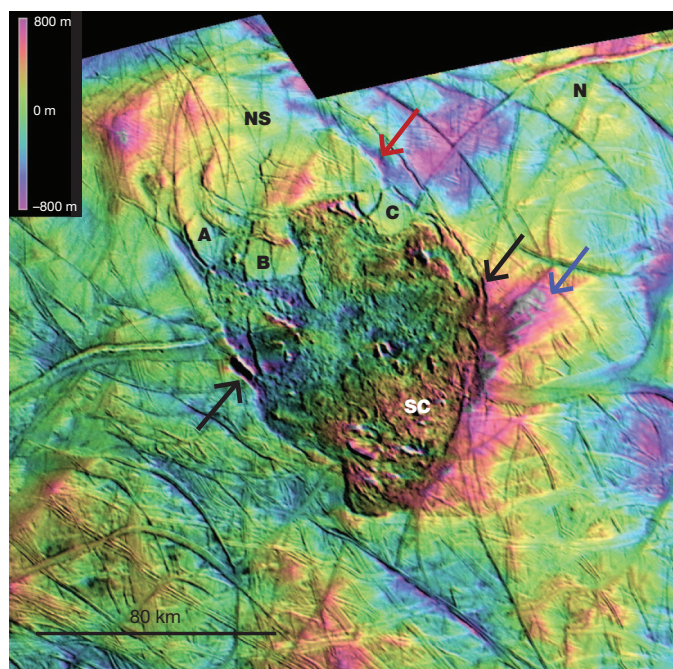


**Figure 1 | Conamara Chaos is dominated by long-wavelength topography.** The region's ( $8^{\circ}$  N,  $274^{\circ}$  W) topography was analysed by filtering a DEM (Digital Elevation Map) produced through a combined stereo photogrammetric and photoclinometric method (Supplementary Information section 1). The DEM has a spatial resolution of 180 m per pixel, and height accuracy of 20 m. Colour indicates topographic heights relative to the background terrain. **a**, Conamara Chaos. The area boxed in white (bottom left) indicates the locations of the top, middle and bottom insets shown right. Insets show, top to bottom, the short-wavelength ( $<20$  km), absolute, and long-wavelength ( $>20$  km) topographic signals produced by the DEM filtering. **b**, **c**, North-south (**b**) and east-west (**c**) topographic profiles of a typical dome-like matrix swell. Solid lines, the DEM; dashed and dashed-dot lines, the long- and short-wavelength topography, respectively. The profiles are offset in elevation for clarity and vertical exaggeration is 40:1. Overall, the topography is characterized by highs within the disaggregated ice matrix and lows at large

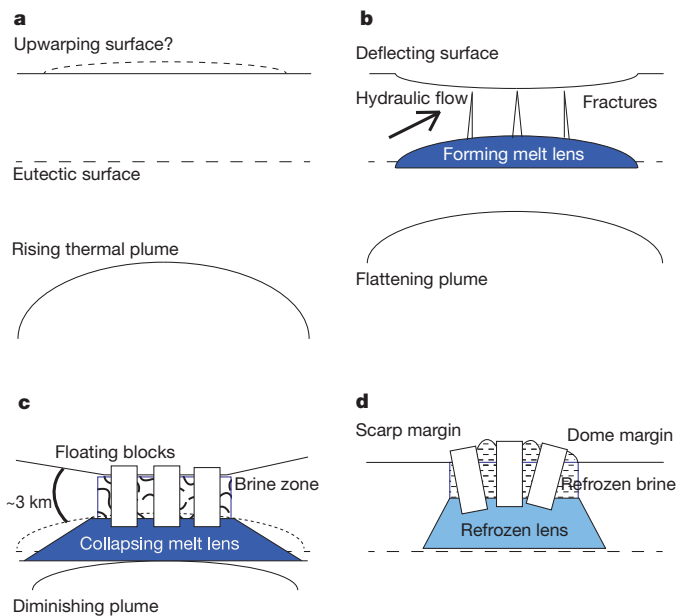
polygonal block-like icebergs. Conamara Chaos appears to have completely disrupted the ice fringed by its boundary scarp, and to have thickened relative to the background terrain. Matrix 'domes' reach heights typically of 200 m. These domes can entrain or tilt some smaller blocks. Large blocks represent the lowest points within the region, with heights equivalent to or up to 100 m below the background elevation. On average, the entire region is raised by  $\sim 100$  m. Comparing insets in **a**, the short-wavelength topography has little spatial variation, while the spatial patterns of the DEM and long-wavelength topography are similar, demonstrating that the long-wavelength signal contributes strongly to Conamara Chaos's heterogeneous topography. That the highest topography is (1) rounded in shape and (2) located within or 'controlled' by matrix swells, indicates that water injection within the matrix and subsequent freezing is responsible for topographic heterogeneities within mature chaos terrain.

be porous. This weak ice can then be disrupted by transient pressure from below, by diurnal tides, by wedging in freezing cracks or by interactions with translating blocks, eventually breaking down into the observed impurity-rich matrix<sup>1,3,7,18</sup>. This mechanism will convert plains material into matrix without requiring thermodynamically unlikely surface melt (see, for example, refs 10, 18).

Conamara Chaos is not only elevated above the background terrain, but also contains matrix domes that are in some places higher than blocks<sup>9,18</sup> (Fig. 1). This observation was initially interpreted as resulting from the coalescence of multiple warm ice diapirs<sup>9</sup>. However, Conamara Chaos's continuity and nearly unbroken margin is more consistent with a single source of disruption. The lens-collapse model implies that brines are injected into former plains material as the ice fractures (Supplementary Fig. 6), while increased water pressure from the lens freezing raises the saturation level within the permeable matrix



**Figure 2 | Thera Macula is a region of likely active chaos production above a large liquid water lens.** The image of Thera Macula was produced using photoclinometry (Supplementary Information section 1) of Galileo images with illumination from the north (N) and 220 m per pixel resolution. Colour indicates topographic heights relative to background terrain. The region ( $50^{\circ}$  S,  $180^{\circ}$  W) exhibits Conamara-like icebergs and dark matrix. The centre of Thera Macula is sunken below the background terrain (denoted here by pale green) by up to  $\sim 800$  m just outside a large semi-circular northern subsiding province (NS), and shows evidence for thickening of matrix in its southern chaotic terrain (SC), which is elevated above the background terrain by up to  $\sim 800$  m. Despite the distinct difference between the morphologies of the NS and the SC, they share a nearly continuous circular scarp boundary (red and black arrows), suggesting that they formed from a common subsurface disruption. Blocks A, B and C are calving at reactivated pre-existing fractures. Within the NS, the platy-blocks appear bent, presumably by basal thinning and subsidence, and the edge of the NS appears ready to break or is being thinned (red arrow). Tall scarps (black arrows) either cast shadows (southern facing) within the region's interior or have bright faces (northern facing), demonstrating the relatively low-lying nature of the centre of Thera Macula. To the lower right, an older band or ridge complex interrupted by the disruption of Thera Macula appears to be swelling along its ridge-like lineations, consistent with brine infiltration and refreeze (blue arrow). The still sunken topography of Thera Macula is indicative of subsurface water.



**Figure 3 | A new hypothesis for chaos formation.** **a**, An ascending thermal plume in the subsurface approaches the pressure-melting eutectic point of the overlying impure brittle ice. **b**, Melting causes surface subsidence that hydraulically confines water, and produces tensile cracks. This behaviour is governed by the relationship<sup>19,20</sup>

$$\nabla\phi_b = (\rho_w - \rho_i)g\nabla z_p + g\rho_i\nabla z_s \quad (1)$$

where  $\phi_b$  is the fluid potential at the lens base (initially the plume head),  $\rho_i$  and  $\rho_w$  are the densities of ice and water, and  $z_p$  and  $z_s$  are the elevations of the base and the surface relative to the geoid, respectively. The slope factor,  $\rho_i/(\rho_w - \rho_i)$ , which is dependent on the impurity content of the ice and water, determines the relative importance of the surface and basal slopes in controlling the direction of water flow. Equation (1) requires that the melt form a lens-shaped pocket with finite effective pressure due to the plume below and overburden from the ice above, unless the slope over the lens is  $\rho_i/(\rho_w - \rho_i)$  times steeper (and in the opposite sense) than the surface slope within the depression. Ultimately the lens top must reach hydraulic equilibrium between the melt volume and the melt-induced depression overlying it. **c**, Hydrofracture from the melt lens calves ice blocks, while fracture and brine infiltration form a granular matrix. **d**, Refreezing of the melt lens and freezing of now brine-rich matrix raises the chaos feature above surrounding terrain, and can cause domes to form between blocks and at margins. Note also that the topography at chaos margins can depend both on the ice type at the margin as well as the direction of the ice collapse. Where deep fractures of blocks define the margin, steep scarps are expected, whereas boundaries defined by matrix (or shallow fractures) will warp into a dome.

(Fig. 3d). This now brash ice eventually swells in response to thermal expansion of freezing brines that fill its once empty pores and cracks. The thickening of brine-infiltrated ice will raise chaos terrains above undisrupted blocks of background terrain and form matrix domes, as confirmed by our analysis of the detailed topography of Conamara Chaos (Fig. 1; Supplementary Information section 3).

Whereas block morphology and motion<sup>3,11</sup> chronicle chaos kinematics, matrix preserves a record of its thermodynamics and Europa's ice rheology. The addition of a minimum of 940 km<sup>3</sup> of liquid water into the matrix is required to produce the average height of Conamara Chaos (Supplementary Information section 3); this is probably only a small fraction of the lens required to produce such a large feature. Also, the stability of any floating ice block to toppling within the matrix depends strongly on its shape<sup>27,28</sup>; rectangular blocks are stable against toppling for aspect ratios (thickness to width) below  $\sim 1.4$  (ref. 28). The smallest tilting blocks at Conamara Chaos thus give us an estimate of the depth to liquid water when they formed. The smallest upright but tilting (conditionally stable) blocks are  $\sim 2$  km wide<sup>11</sup>, and thus  $\sim 2.8$  km thick, providing an estimate of the brittle

layer's thickness and the minimum depth of water lenses. Ice block and matrix dome heights within Conamara Chaos (Fig. 1) are consistent with 10–35% ice porosity for reasonable brine densities (Supplementary Information section 3).

Most importantly, the lens-collapse model developed here makes testable predictions: whereas swelling matrix indicates lens freeze-out, liquid water may be found where the surface subsides and blocks 'float' above the matrix. At Thera Macula, we are probably witnessing active chaos formation (Fig. 2). The large concentric fracture system encircling Thera Macula resembles those of collapsing ice cauldrons<sup>19,20</sup> (Supplementary Fig. 4), and, given the absence of a continuous moat, suggests that subsurface melt and ice disaggregation is forming Thera Macula, rather than the collapse of a dome<sup>29</sup>. Topographic contrast indicates that the rapid freeze of matrix is occurring at the region's south, but surface slopes, ice blocks and incomplete break-up to the north indicates that the lens below Thera Macula was liquid at the time of the Galileo encounter. Today, a melt lens of 20,000–60,000 km<sup>3</sup> of liquid water probably lies below Thera Macula; this equates to at least the estimated combined volume of the Great Lakes (Supplementary Information section 3). Although it is unclear how rapidly the break-up of Thera Macula took place, such a volume would take  $\sim 10^5$ – $10^6$  years to freeze. Surface modification should be extensive just after the collapse and persist as long as the lens is mostly liquid, such that Thera Macula may have noticeable changes between the Galileo encounter and the present day. Such features can be well understood by coupled topographic and subsurface imaging<sup>30</sup>.

The existence of globally distributed chaos terrain<sup>1,2</sup> argues that pervasive shallow subsurface water has existed and continues to exist within Europa's icy shell; surface draw down at Thera Macula suggests that it exists within 3 km of the surface. The lens-collapse model presented here explains previously discrepant observations of chaos (refs 10, 18; Supplementary Information section 4). Our work predicts that it is the scale of ascending plumes and local surface geology that produces diverse chaos morphologies, and thus our model may be extended to other features, such as Murias Chaos, as well as pits and domes. Our analyses suggest that ice–water dynamics are active today on Europa, sustaining large liquid lakes perched in the shallow subsurface.

Received 17 April; accepted 4 October 2011.

Published online 16 November 2011.

- Pappalardo, R. *et al.* Geological evidence for solid-state convection in Europa's ice shell. *Nature* **391**, 365–368 (1998).
- Figueredo, P. H. & Greeley, R. Resurfacing history of Europa from pole-to-pole geological mapping. *Icarus* **167**, 287–312 (2004).
- Carr, M. H. *et al.* Evidence for a subsurface ocean on Europa. *Nature* **391**, 363–365 (1998).
- Squyres, S. W., Reynolds, R. T., Cassen, P. & Peale, S. J. Liquid water and active resurfacing on Europa. *Nature* **301**, 225–226 (1983).
- Cassen, P., Reynolds, R. T. & Peale, S. J. Is there liquid water on Europa? *Geophys. Res. Lett.* **6**, 731–734 (1979).
- Kivelson, M. G. *et al.* Galileo magnetometer measurements: a stronger case for a subsurface ocean at Europa. *Science* **289**, 1340–1343 (2000).
- Greenberg, R. G. *et al.* Chaos on Europa. *Icarus* **141**, 263–286 (1999).
- O'Brien, D. P., Geissler, P. & Greenberg, R. A melt through model for chaos formation on Europa. *Icarus* **156**, 152–161 (2002).
- Schenk, P. & Pappalardo, R. T. Topographic variations in chaos on Europa: implications for diapiric formation. *Geophys. Res. Lett.* **31**, L16703, doi:10.1029/2004GL019978 (2004).
- Collins, G. C., Head, J. W. III, Pappalardo, R. T. & Spaun, N. A. Evaluation of models for the formation of chaotic terrain on Europa. *J. Geophys. Res.* **105**, 1709–1716 (2000).
- Spaun, N. A. *et al.* Conamara Chaos region, Europa: reconstruction of mobile polygonal ice blocks. *Geophys. Res. Lett.* **25**, 4277–4280 (1998).
- Rathbun, J. A., Musser, G. S. Jr & Squyres, S. W. Ice diapirs on Europa: implications for liquid water. *Geophys. Res. Lett.* **25**, 4157–4160 (1998).
- Pappalardo, R. & Barr, A. C. The origin of domes on Europa: the role of thermally induced compositional diapirism. *Geophys. Res. Lett.* **31**, L01701, doi:10.1029/2003GL019202 (2004).
- Head, J. W. & Pappalardo, R. T. Brine mobilization during lithospheric heating on Europa: implications for formation of chaos terrain, lenticular texture, and color variations. *J. Geophys. Res.* **104** (E11), 27143–27155 (1999).



# Structural transformation in supercooled water controls the crystallization rate of ice

Emily B. Moore<sup>1</sup> & Valeria Molinero<sup>1</sup>

One of water's unsolved puzzles is the question of what determines the lowest temperature to which it can be cooled before freezing to ice. The supercooled liquid has been probed experimentally to near the homogeneous nucleation temperature,  $T_H \approx 232$  K, yet the mechanism of ice crystallization—including the size and structure of critical nuclei—has not yet been resolved. The heat capacity and compressibility of liquid water anomalously increase on moving into the supercooled region, according to power laws that would diverge (that is, approach infinity) at  $\sim 225$  K (refs 1, 2), so there may be a link between water's thermodynamic anomalies and the crystallization rate of ice. But probing this link is challenging because fast crystallization prevents experimental studies of the liquid below  $T_H$ . And although atomistic studies have captured water crystallization<sup>3</sup>, high computational costs have so far prevented an assessment of the rates and mechanism involved. Here we report coarse-grained molecular simulations with the mW water model<sup>4</sup> in the supercooled regime around  $T_H$  which reveal that a sharp increase in the fraction of four-coordinated molecules in supercooled liquid water explains its anomalous thermodynamics and also controls the rate and mechanisms of ice formation. The results of the simulations and classical nucleation theory using experimental data suggest that the crystallization rate of water reaches a maximum around 225 K, below which ice nuclei form faster than liquid water can equilibrate. This implies a lower limit of metastability of liquid water just below  $T_H$  and well above its glass transition temperature, 136 K. By establishing a relationship between the structural transformation in liquid water and its anomalous thermodynamics and crystallization rate, our findings also provide mechanistic insight into the observed<sup>5</sup> dependence of homogeneous ice nucleation rates on the thermodynamics of water.

We performed large-scale molecular dynamics simulations with the mW water model, which represents a water molecule as a single particle with short-range anisotropic interactions that mimic hydrogen bonds<sup>4</sup>, to determine the temperature dependence of key thermodynamic properties of liquid water. Figure 1a presents the enthalpy, the heat capacity and the excess free energy with respect to ice, as liquid water is cooled at the lowest rate that still produces low-density amorphous ice (LDA) in simulations with the mW water model. We note that in our simulations most molecules in LDA are four-coordinated, as in ice albeit without long-range order<sup>6,7</sup>, and that the structural transformation from liquid water to amorphous ice is sharp but continuous. This sharp yet continuous nature of the structural transition agrees with the conclusions from experiments on water confined in narrow silica nanopores that prevent ice crystallization<sup>8</sup> and from a thermodynamic analysis<sup>9</sup> of bulk water outside the so-called 'no-man's land' that stretches between the glass transition temperature ( $T_g$ ) and  $T_H$ .

The enthalpy of liquid water (Fig. 1a, top panel) decreases steeply around the liquid transformation temperature  $T_L = 202 \pm 2$  K (defined by the maximum change in density) and approaches the value for ice (Supplementary Fig. 1). The heat capacity  $C_p$  (Fig. 1a, middle panel) reaches a maximum at  $T_L$ , which is also the locus of maximum change

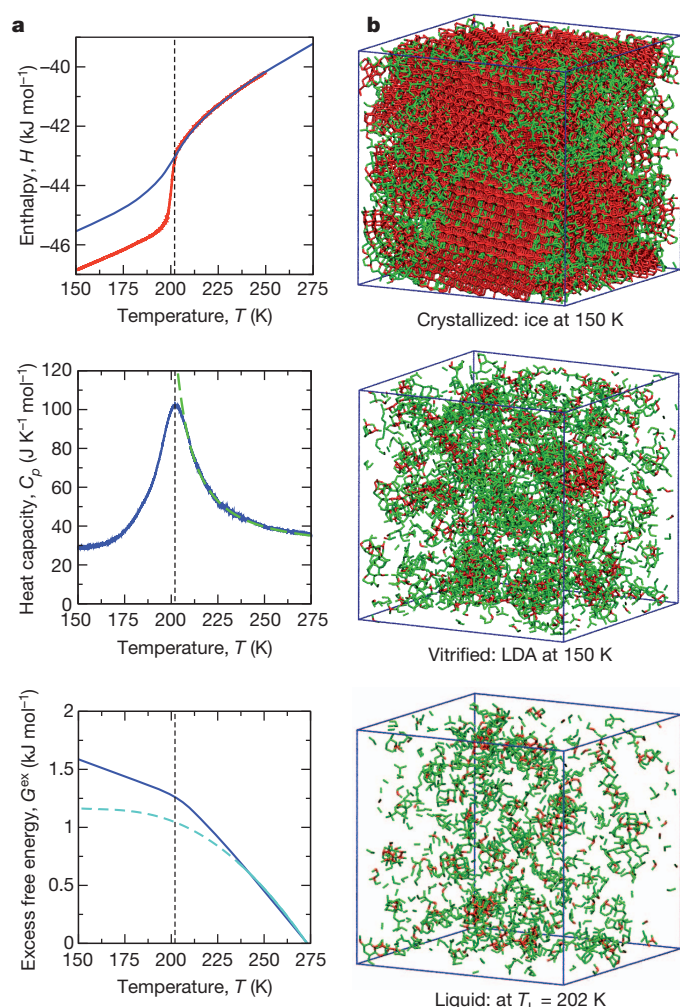
in tetrahedrality and fraction of four-coordinated molecules<sup>7</sup>. We note that  $T_L$  in the simulations is  $\sim 15$  K above the singular temperature of the power law,  $T_s$ , predicted by a fit of  $C_p$  obtained with the mW water<sup>4</sup>, and  $\sim 25$  K below the  $T_s \approx 225$  K, estimated from the experimental  $C_p$  of water<sup>1,2</sup>. Large patches of four-coordinated molecules—the signature of LDA, and in an earlier simulation<sup>3</sup> identified as the precursor of the ice nucleus—develop in supercooled water and grow on cooling following a power law that would peak at  $T_L$  (ref. 7); recent small angle X-ray scattering experiments on supercooled water down to 250 K (ref. 10) concur with our simulation results.

The development of crystallinity in water on cooling is illustrated in Fig. 1b. We note that calorimetry and X-ray diffraction report 5% ice in LDA<sup>11</sup>, which is the same fraction we find in our simulations (Supplementary Fig. 2). We also find that ice in LDA appears as small crystallites surrounded by threads of water with a local structure intermediate between ice and the four-coordinated liquid, and without long-range order. This 'intermediate-ice' structure accounts for  $\sim 20\%$  of all the water present in LDA (Supplementary Fig. 2) and may be a realization of the 'gossamer percolative network' of nanocrystallites predicted to form at temperatures for which the length scale for motions relevant to the structural relaxation of the liquid is larger than the critical nucleus size for crystallization<sup>12</sup>. The large number of threads of intermediate-ice that appear on approaching and crossing  $T_L$  illustrates a blurring of the boundary between clearly liquid structures and clearly crystalline structures in deeply supercooled water. Below we show that this blurring heralds the effective limit of metastability of liquid water.

Experimentally observed crystallization rates increase when cooling liquid water towards  $T_H$  but increase on heating the glass around  $T_g$ , implying the existence of a temperature  $T_x$  of maximum crystallization rate in water's 'no-man's land'. The top panel of Fig. 1a indicates that on decreasing the rate of cooling, crystallization—evidenced by a sharp decrease in enthalpy—first occurs at  $T_L$  and should thus be fastest at that temperature. Figure 2a presents the time-temperature-transformation (TTT) diagram of mW water; the circles indicate the ice crystallization times,  $\tau_x$ , computed from more than 1,000 independent simulations. The data show that  $\tau_x$  is minimum at  $T_x \approx 200$  K, almost identical to  $T_L = 202 \pm 2$  K. For comparison, Fig. 2b shows the TTT curve obtained when using nucleation theory and experimental data for water (Supplementary discussion A): the crystallization time  $\tau_x$  is predicted to be minimum and the crystallization rate maximum at  $T_x \approx 225$  K, close to  $T_s$  and just a few degrees below  $T_H$ .

To disentangle the contributions of structural transformation and degree of supercooling in determining  $T_x$ , we investigated the freezing of water that is confined in a 3-nm-diameter cylindrical nanopore and therefore exhibits a decrease in the melting temperature of ice from 273 to 220 K for both the mW model<sup>13</sup> and experiment<sup>14</sup>. The simulations indicate that  $T_L \approx T_x \approx 200$  K in the pore (Supplementary Fig. 3). We conclude that the freezing temperature of water is controlled by the structural transformation of the liquid and not merely the degree of supercooling. This explains the experimentally observed closing of the

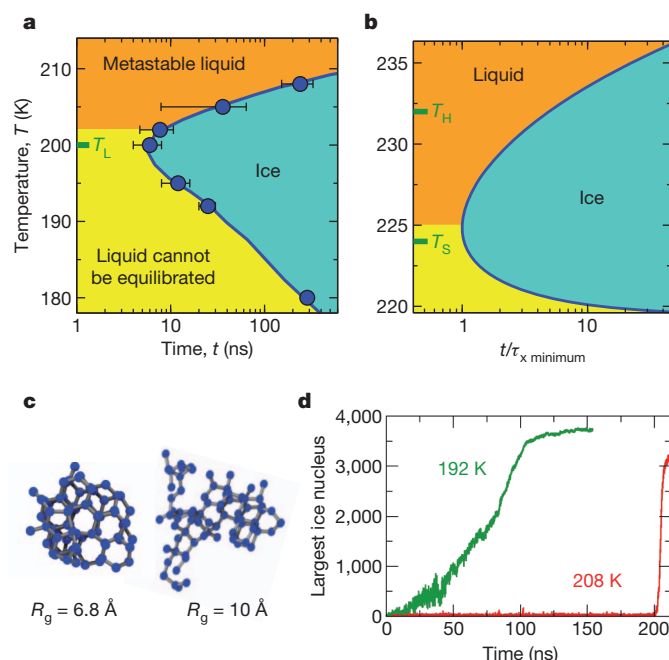
<sup>1</sup>Department of Chemistry, University of Utah, Salt Lake City, Utah 84112-0580, USA.



**Figure 1 | Evolution of the thermodynamics and structure of water on cooling.** **a**, Enthalpy  $H$  (top panel), heat capacity  $C_p$  (middle) and excess free energy  $G^{\text{ex}}$  (bottom) of liquid mW water on hyperquenching to LDA glass at  $10 \text{ K ns}^{-1}$  (blue solid lines). Vertical dashed line indicates the liquid transformation temperature  $T_L$ , which sets a lower limit of metastability (LLM) of liquid water. Thus  $H$ ,  $C_p$  and  $G^{\text{ex}}$  for  $T < T_L$  are not equilibrium quantities and depend on the cooling rate. Cooling mW water at  $1 \text{ K ns}^{-1}$  (red line, top) results in crystallization at  $T_L$  where the crystallization rate is maximum. On cooling towards  $T_L$ ,  $C_p$  follows a power law (green dashed line, middle)<sup>4</sup> with exponent 1.5 as in experiments<sup>2</sup> and  $T_s$  about 35 K lower.  $G^{\text{ex}}$  in the simulations (blue line, bottom) is in excellent agreement with  $G^{\text{ex}}$  in experiments (solid cyan line)<sup>9</sup>; the dashed cyan line shows the experimental  $G^{\text{ex}}$  extrapolated below  $T_H$  (ref. 5). **b**, Ice (red) and intermediate-ice (green) in ice formed by cooling water at  $1 \text{ K ns}^{-1}$  (top), LDA formed by quenching water at  $10 \text{ K ns}^{-1}$  (middle) and in liquid water at  $T_L$  (bottom). Lines connect water molecules within  $3.5 \text{ \AA}$ .

gap between freezing and melting of confined water on decreasing the radius of the confining nanopore<sup>14</sup>.

We also determined the number of water molecules  $N^*$  in the critical ice nuclei (including their sheath of intermediate-ice) through the mean first passage time (MFPT) method<sup>15</sup>. Our values of  $N^*$  ( $\sim 120$  at  $T_L + 6 \text{ K}$  and  $\sim 90$  at  $T_L + 3 \text{ K}$ ) are in good agreement with  $N^*$  (70–210) deduced from freezing of water in micelles around  $T_H$  (ref. 16). The critical nucleus size is determined by liquid–ice thermodynamics, well reproduced by the mW model (Supplementary discussions B and C). Critical nuclei around  $T_L + 3 \text{ K}$  have a broad distribution of shapes (Fig. 2c), indicating a lowering of the ice–liquid surface tension as the structural gap between ice and liquid narrows on approaching  $T_L$ . The fraction of four-coordinated water molecules in the liquid wetting the nuclei at  $T_L + 3 \text{ K}$  is 50% larger than the average for the whole liquid (Supplementary Fig. 4); this corroborates the conjecture put forward in



**Figure 2 | Kinetics of ice crystallization and critical ice nuclei.** **a**, Time–temperature transformation (TTT) diagram of mW water. Blue circles indicate average times  $\tau_x$  to crystallize 70% of water. The error bars indicate the range of crystallization times measured in the simulations and show a large dispersion of crystallization times above  $T_L$  due to the stochastic nature of the nucleation process. Water crystallizes within the ‘nose’ (the blue-shaded area). The crystallization time is the shortest around 200 K, close to  $T_L = 202 \pm 2 \text{ K}$ . Nucleation and growth times become comparable at  $T_L$  (Supplementary Fig. 5). Above  $T_L$ , crystallization is limited by the rare formation of critical ice nuclei and supercooled water can be studied in the metastable liquid state (orange shaded area). Below  $T_L$ , crystallization occurs before relaxation of liquid water and the liquid exists only for times too short for its equilibration (yellow shaded area). The maximum crystallization rate predicted by the mW model is  $J = (\tau_x V)^{-1} \approx 10^{27} \text{ cm}^{-3} \text{ s}^{-1}$ , several orders of magnitude faster than measured for water down to  $T_H$ , because crystallization rates are proportional to water mobility (Supplementary Information section A), which is overestimated by the mW model<sup>4,27</sup>. **b**, TTT curve predicted using classical nucleation theory and experimental data for water. The x axis represents the time normalized by the minimum value of the crystallization time,  $\tau_{x,\text{minimum}}$ . The nose resulting from crossover between nucleation and growth occurs at 225 K, between the experimental  $T_H$  and  $T_s$ . The colour code of the shaded areas is the same as in **a**. **c**, Critical ice nuclei at the lowest temperature for which liquid water can be equilibrated,  $T_L + 3 = 205 \text{ K}$  in the simulations, contain  $\sim 90$  water molecules and a wide range of compactness. The spread in the radius of gyration ( $R_g$ ) of the nuclei (Supplementary Figs 6 and 7) suggests that the liquid–ice surface tension is very low on approaching  $T_L$ . **d**, Number of molecules in the largest ice nucleus for representative simulations at 208 K (red line; at this temperature, crystallization is dominated by stochastic nucleation) and at 192 K (green line; where nucleation is fast and crystallization proceeds at the pace of growth).

an earlier study<sup>3</sup> and suggests that four-coordinated water patches that form in the supercooled liquid stabilize the crystal nuclei.

The minimum in crystallization times around  $T_L$  signals a crossover in the mechanism of ice crystallization, from nucleation-dominated above  $T_L$  to growth-dominated below  $T_L$  (Fig. 2d). The lack of a well-defined nucleation plateau in the MFPT plot at  $T_L$  (Supplementary Fig. 5) is evidence of concurrent nucleation and growth, and implies that the barrier for nucleation is comparable to the thermal energy,  $RT$  (where  $R$  is the gas constant; Supplementary discussion D)<sup>15</sup>. The growth time of the crystallites is comparable to or shorter than the relaxation time of the liquid because the growth rate is proportional to the diffusivity of supercooled liquid water<sup>17</sup>, which decouples from the structural relaxation<sup>18</sup> (Supplementary discussions A and D). Thus liquid water cannot be equilibrated in the simulations at  $T < T_L + 3 \text{ K}$ : ice nucleates before the liquid has time to equilibrate. We conclude that



the structural transformation in supercooled water around  $T_L$  sets the effective lower limit of metastability (LLM) of supercooled water.

In fact, a kinetic limit of stability for the liquid state was anticipated by Kauzmann<sup>19</sup>, as the resolution of his entropy paradox: “the barrier to crystal nucleus formation, which tends to be very large just below the melting point, may at low temperatures be reduced to approximately the same height as the free energy barriers which impede molecular reorientations in the liquid (...). Under these circumstances crystal nuclei will form and grow at about the same rate as the liquid changes its structure following a change in temperature or pressure.” An extension of this argument considered the decoupling of diffusion and viscosity in supercooled liquids and concluded that an LLM must be reached at  $T_{LLM} > T_o$ , where the latter indicates the temperature at which the excess entropy of the liquid would become lower than the entropy of the crystal<sup>17</sup>. The experimental power-law increase in  $C_p$  of liquid water on cooling supports the existence of an LLM of water between  $T_{LLM} > T_o \approx T_s \approx 225$  K and  $T_H \approx 232$  K (Supplementary discussions A and D); our analysis concurs with the prediction of a kinetic spinodal in water at  $\sim 230$  K based on the fluctuation theory of relaxation of metastable states<sup>20</sup>.

The above arguments for real water, using classical nucleation theory and experimental data, predict a lower limit of metastability close to the ‘nose’ of the TTT curve, between  $T_s$  and  $T_H$ —as predicted by our simulations. We therefore conclude that between  $T_{LLM}$  ( $\approx T_L > T_s$ ) and  $T_g$ , liquid water is not metastable and can only be studied over times shorter than needed for its equilibration. The low nucleation barriers and considerable water diffusivity around  $T_{LLM}$  make partial crystallization unavoidable, even at the fastest attainable cooling rates. Therefore the glass transition of LDA at  $T_g \approx 136$  K does not produce metastable liquid water, but rather a less viscous liquid unable to relax before crystallizing. The lack of ergodicity in liquid water below  $\sim 225$  K may explain the feeble heat capacity signature at  $T_g$  that puzzled scientists for decades<sup>21</sup>.

Various theoretical scenarios, involving a retracing spinodal of superheated water, a first-order and a continuous liquid–liquid transition, have been proposed to explain the thermodynamic anomalies of water and predict its fate in ‘no-man’s land’<sup>22–25</sup>. These scenarios assume that metastable liquid water exists below  $T_H$ . The results of the present work suggest that the structural transformation that causes the anomalies of water is also responsible for the demise of the liquid state. Water has been proposed to first convert to low-density liquid (LDL) and then crystallize<sup>26</sup>, but our results (reported here and in ref. 27) suggest that crystallization occurs faster than LDL’s equilibration. It has been argued that LDL can be equilibrated in simulations using the ST2 model<sup>23,28</sup>, recently reported free energy maps of ST2 and mW models, however, do not display a basin for LDL<sup>29</sup>. We also note that our calculations indicate that water crystallization in ‘no-man’s land’ is limited only by the growth rate of the crystallites, which decreases on cooling. So extrapolation of crystallization rates from the nucleation-dominated region above  $T_H$  to temperatures below 225 K—that is, to temperatures relevant for cloud formation and crucial for the formulation of climate models—would severely overestimate the rates of ice formation.

## METHODS SUMMARY

Simulations were performed with LAMMPS<sup>30</sup> using the mW water model<sup>4</sup>, which reproduces the structure, anomalies and phase behaviour of water at less than 1% of the computational cost of atomistic models (Supplementary discussion B). Thermodynamic properties of the liquid were computed as indicated in Online Methods for simulation cells containing 32,768 molecules at pressure  $p = 1$  atm and a linear decrease in temperature at  $10\text{ K ns}^{-1}$ . Ice was identified with the CHILL algorithm<sup>13</sup>. Crystallization times indicate average times to crystallize 70% of  $\sim 150$  independent constant pressure and temperature simulations with 4,096 molecules at each temperature. Nucleation times and the average critical size and radius of gyration of the nuclei were determined from MFPT analysis<sup>15</sup> of the crystallization trajectories. Identification of individual critical nuclei was performed through evaluation of their individual crystallization probability over 200 independent simulations for each nucleus, starting with the same configuration and randomized momenta.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 4 February; accepted 13 September 2011.**

- Speedy, R. J. & Angell, C. A. Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at  $-45^\circ\text{C}$ . *J. Chem. Phys.* **65**, 851–858 (1976).
- Tombari, E., Ferrari, C. & Salvetti, G. Heat capacity anomaly in a large sample of supercooled water. *Chem. Phys. Lett.* **300**, 749–751 (1999).
- Matsumoto, M., Saito, S. & Ohmine, I. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature* **416**, 409–413 (2002).
- Molinero, V. & Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **113**, 4008–4016 (2009).
- Koop, T., Luo, B. P., Tsias, A. & Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **406**, 611–614 (2000).
- Finney, J. L., Hallbrucker, A., Kohl, I., Soper, A. K. & Bowron, D. T. Structures of high and low density amorphous ice by neutron diffraction. *Phys. Rev. Lett.* **88**, 225503 (2002).
- Moore, E. B. & Molinero, V. Growing correlation length in supercooled water. *J. Chem. Phys.* **130**, 244505 (2009).
- Mallamace, F. *et al.* The anomalous behavior of the density of water in the range  $30\text{ K} < T < 373\text{ K}$ . *Proc. Natl Acad. Sci. USA* **104**, 18387–18391 (2007).
- Johari, G. P., Fleissner, G., Hallbrucker, A. & Mayer, E. Thermodynamic continuity between glassy and normal water. *J. Phys. Chem.* **98**, 4719–4725 (1994).
- Huang, C. *et al.* Increasing correlation length in bulk supercooled H<sub>2</sub>O, D<sub>2</sub>O, and NaCl solution determined from small angle x-ray scattering. *J. Chem. Phys.* **133**, 134504 (2010).
- Kohl, I., Mayer, E. & Hallbrucker, A. The glass water-cubic ice system: a comparative study by X-ray diffraction and differential scanning calorimetry. *Phys. Chem. Chem. Phys.* **2**, 1579–1586 (2000).
- Stevenson, J. D. & Wolynes, P. G. The ultimate fate of supercooled liquids. *J. Phys. Chem. A* **115**, 3713–3719 (2011).
- Moore, E. B., de la Llave, E., Welke, K., Scherlis, D. A. & Molinero, V. Freezing, melting and structure of ice in a hydrophilic nanopore. *Phys. Chem. Chem. Phys.* **12**, 4124–4134 (2010).
- Jähnert, S. *et al.* Melting and freezing of water in cylindrical silica nanopores. *Phys. Chem. Chem. Phys.* **10**, 6039–6051 (2008).
- Wedekind, J., Strey, R. & Reguera, D. New method to analyze simulations of activated processes. *J. Chem. Phys.* **126**, 134103 (2007).
- Liu, J., Nicholson, C. E. & Cooper, S. J. Direct measurement of critical nucleus size in confined volumes. *Langmuir* **23**, 7286–7292 (2007).
- Tanaka, H. Possible resolution of the Kauzmann paradox in supercooled liquids. *Phys. Rev. E* **68**, 011505 (2003).
- Xu, L. *et al.* Appearance of a fractional Stokes-Einstein relation in water and a structural interpretation of its onset. *Nature Phys.* **5**, 565–569 (2009).
- Kauzmann, W. The nature of the glassy state and the behavior of liquids at low temperatures. *Chem. Rev.* **43**, 219–256 (1948).
- Kiselev, S. Physical limit of stability in supercooled liquids. *Int. J. Thermophys.* **22**, 1421–1433 (2001).
- Angell, C. A. Insights into phases of liquid water from study of its unusual glass-forming properties. *Science* **319**, 582–587 (2008).
- Speedy, R. J. Stability-limit conjecture — an interpretation of the properties of water. *J. Phys. Chem.* **86**, 982–991 (1982).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
- Sastry, S., Debenedetti, P. G. & Sciortino, F. Singularity-free interpretation of the thermodynamics of supercooled water. *Phys. Rev. B* **53**, 6144–6154 (1996).
- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
- Mishima, O. Application of polyamorphism in water to spontaneous crystallization of emulsified LiCl–H<sub>2</sub>O solution. *J. Chem. Phys.* **123**, 154506 (2005).
- Moore, E. B. & Molinero, V. Ice crystallization in water’s “no-man’s land”. *J. Chem. Phys.* **132**, 244504 (2010).
- Liu, Y., Panagiotopoulos, A. Z. & Debenedetti, P. G. Low-temperature fluid-phase behavior of ST2 water. *J. Chem. Phys.* **131**, 104508 (2009).
- Limmer, D. T. & Chandler, D. The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. *J. Chem. Phys.* **135**, 134503 (2011).
- Plimpton, S. J. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Arnold and Mabel Beckman Foundation through a Young Investigator Award to V.M. We thank P. G. Debenedetti for discussions and D. P. Fernandez for criticism of the manuscript.

**Author Contributions** V.M. conceived and designed the study and wrote the paper. E.B.M. and V.M. performed the simulations, analysed the data and interpreted the results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to V.M. (Valeria.Molinero@utah.edu).

## METHODS

**Simulations.** All molecular dynamics simulations were performed with LAMMPS<sup>30</sup>. Equations of motion were integrated using Velocity Verlet with a time step of 10 fs. Bulk simulations were conducted in the  $NpT$  ensemble, with  $p = 1$  atm. Temperature and pressure were controlled with the Nose-Hoover thermostat and barostat, with time constants of 1 and 5 ps, respectively. The target temperature was decreased linearly in the cooling ramp simulations. Water was modelled with the mW potential<sup>4</sup>. Three different systems were used in this study: (1) the thermodynamics of bulk water was determined through cooling ramps with simulation cells containing 32,768 mW water molecules; (2) the isothermal crystallization simulations were performed with cells containing 4,096 mW water molecules, after checking that it produced consistent results with a simulation cell containing 13,768 molecules; and (3) the confined water system consisted of the 3 nm cylindrical nanopore used in the studies of ref. 31 and contained 2,123 water molecules embedded in a 5,840-molecule pore. The interactions between pore-wall and water are chosen to be as water with water, to minimize the effect of the pore-wall on the liquid<sup>13</sup>. The pore was headless to ensure that the crystallization occurs within the shaft of the pore. The pore was 90% filled to allow for expansion of the water as it is cooled and forms ice. The nanopore simulations were performed in the  $NVT$  ensemble, although it should be noted that the water inside is at zero pressure as the pore is not fully filled.

**Thermodynamic properties.** The enthalpy,  $H = E + pV$ , was directly computed along the simulation trajectories and saved every 0.2 ps or less and averaged over 100 ps running intervals. The heat capacity was obtained through numerical differentiation of  $H$  with respect to temperature. The excess entropy of liquid water with respect to ice (Supplementary Fig. 1) was obtained through integration of the change in entropy from the value at the melting point,  $\Delta S_m = \Delta H_m/T_m$ ,

$$S^{\text{ex}}(T) = \Delta S_m(T_m) - \int_{T_m}^T \frac{C_p^{\text{ex}}(T')}{T'} dT'. \quad (1)$$

where the excess heat capacity of liquid with respect to ice is  $C_p^{\text{ex}}(T') = C_p^{\text{liquid}}(T') - C_p^{\text{ice}}(T')$ .  $C_p^{\text{ice}}(T)$  was computed in ref. 32, as well as the excess free energy,  $G^{\text{ex}} = H^{\text{ex}} - TS^{\text{ex}}$ , that here we extend down to 150 K. We note that at temperatures lower than 205 K the liquid cannot be equilibrated and the thermodynamic properties depend on the rate of cooling, which determines the fraction of water that crystallizes to ice.

**Identification of the ice nuclei.** The CHILL algorithm<sup>13</sup> was used to distinguish between molecules with local order of liquid, ice I, and molecules with local ordering intermediate between that of ice and liquid, that here we call intermediate-ice and that we have called interfacial ice elsewhere<sup>13,27</sup>, because it also forms on the interface between well defined crystallites and the liquid phase. An ice nucleus consists of clusters of molecules with any ice-like local environment, including both ice I and intermediate-ice. Ice nuclei are defined by clustering of ice and intermediate-ice molecules using 3.5 Å cut-off to define connected neighbours.

**Crystallization simulations at constant temperature.** To produce a large set (more than 100) of independent trajectories at each temperature, starting

configurations were selected at 500 ps intervals from a single simulation at 300 K. From the starting configuration, the temperature was instantaneously quenched to the temperature of interest,  $T_{\text{quench}}$ , from 192 to 208 K, and the time was set to zero. The crystallization time,  $\tau_x$ , is the time required to convert 70% of the water into ice. Nearly 1,000 simulations up to 350 ns in length each were collected. The crystallization time of mW water at 180 K was taken from ref. 27.

**Nucleation times and critical nuclei size.** We used the mean first passage time (MFPT) method as implemented in ref. 15 to determine the characteristic time-scale of nucleation. The number of water molecules in the largest ice nucleus,  $N$ , and its radius of gyration,  $R_g$ , were chosen as order parameters for the advance of the crystallization. The radius of gyration,  $R_g$ , and non-sphericity, NS, of the nuclei were determined as described in ref. 7. With the size of the ice nuclei as the order parameter, the nucleation time  $\tau_{\text{nucleation}}$  and the critical nuclei size,  $N^*$ , can be determined. For a series of trajectories at a given temperature, the mean time of first appearance is recorded for the largest nucleus in each configuration. A plot of the MFPT, the time it takes for a given nucleus size,  $N$ , to grow rather than dissolve for the first time, versus nuclei size results in a sigmoidal curve that can be described by the following equation:<sup>15</sup>

$$\tau(N) = \frac{\tau_{\text{nucleation}}}{2} \{1 + \text{erf}[(N - N^*)c]\} \quad (2)$$

where  $\tau(N)$  is the MFPT as a function of cluster size,  $N$ , and  $c$  is a constant. The plateau of the sigmoidal curve gives the nucleation time, and the inflection point corresponds to the critical nucleus size,  $N^*$ .

**Crystallization probability of individual nuclei.** A series of simulations were run to compare the growth probabilities of selected nuclei of size predicted to be critical by the MFPT method based on radius of gyration. Configurations containing a potentially critically sized nucleus, with size  $N^*$ , were chosen and the  $R_g$  and NS of the nucleus were recorded. A series of 200 simulations were run from a set of independent configurations at 205 K, each initiated with newly randomized velocities, resulting in 200 unique simulations from each starting configuration. The probability of growth from the initial nucleus was calculated as the fraction of trajectories that resulted in nuclei growth after 5 ns, larger than the average growth time at 205 K,  $2 \pm 1$  ns.

**Local liquid environment of the nuclei.** The liquid solvation shell of the crystalline clusters was analysed for the crystallization trajectories at 200 and 205 K. The shell was defined as the molecules of the liquid within 3.5 Å of any molecule of the crystal nucleus (the latter includes the intermediate-ice). We computed the ratio of four-coordinated ( $N_4$ ) to higher-coordinated molecules ( $N_{\text{H}}$ ) around the ice nuclei and compare it with the ratio for the whole system (which is about 40 times larger than the nuclei).

31. de La Llave, E., Molinero, V. & Scherlis, D. A. Water filling of hydrophilic nanopores. *J. Chem. Phys.* **133**, 034513 (2010).
32. Jacobson, L. C., Hujo, W. & Molinero, V. Thermodynamic stability and growth of guest-free clathrate hydrates: a low-density crystal phase of water. *J. Phys. Chem. B* **113**, 10298–10307 (2009).



# Reconstructed changes in Arctic sea ice over the past 1,450 years

Christophe Kinnard<sup>1</sup>, Christian M. Zdanowicz<sup>2</sup>, David A. Fisher<sup>2</sup>, Elisabeth Isaksson<sup>3</sup>, Anne de Vernal<sup>4</sup> & Lonnie G. Thompson<sup>5</sup>

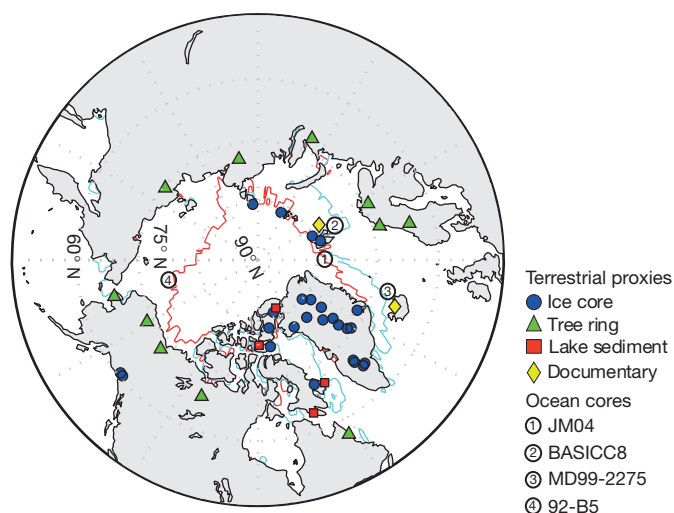
Arctic sea ice extent is now more than two million square kilometres less than it was in the late twentieth century, with important consequences for the climate, the ocean and traditional lifestyles in the Arctic<sup>1,2</sup>. Although observations show a more or less continuous decline for the past four or five decades<sup>3,4</sup>, there are few long-term records with which to assess natural sea ice variability. Until now, the question of whether or not current trends are potentially anomalous<sup>5</sup> has therefore remained unanswerable. Here we use a network of high-resolution terrestrial proxies from the circum-Arctic region to reconstruct past extents of summer sea ice, and show that—although extensive uncertainties remain, especially before the sixteenth century—both the duration and magnitude of the current decline in sea ice seem to be unprecedented for the past 1,450 years. Enhanced advection of warm Atlantic water to the Arctic<sup>6</sup> seems to be the main factor driving the decline of sea ice extent on multidecadal timescales, and may result from nonlinear feedbacks between sea ice and the Atlantic meridional overturning circulation. These results reinforce the assertion that sea ice is an active component of Arctic climate variability and that the recent decrease in summer Arctic sea ice is consistent with anthropogenically forced warming.

Sea ice cover in the Arctic has been steadily decreasing over recent decades, with record losses in recent years<sup>1,2</sup>. The late-summer ice cover, mainly composed of thick multi-year ice, has been shrinking the fastest (−8.6% per decade)<sup>4</sup>, which may soon result in an ice-free Arctic Ocean<sup>7</sup>. The loss of sea ice is having, and will continue to have, profound repercussions on the climate, ocean circulation, ecology and economics of the Arctic<sup>2</sup>. The short period of historical and satellite observations (about 100 years) hampers our understanding of natural variability of sea ice, so that longer time series are much needed to place recent trends in a long-term context. Although a few multi-centennial records exist in the European Arctic<sup>5</sup>, full hemispheric coverage began only in the late nineteenth century<sup>3</sup>. Here we present a reconstruction of variability in sea ice cover in the Arctic over the past 1,450 years based on a network of high-resolution climate proxy records from the circumpolar region. Our proxy-based reconstruction is compared with other temperature and atmospheric circulation reconstructions and indices to identify natural drivers of variability in sea ice in pre-industrial times.

Sea ice cover is thermodynamically and dynamically controlled by both the atmosphere and the ocean. A proxy-based reconstruction should therefore include atmospheric as well as oceanic proxies (such as water temperature or salinity). Unfortunately there are currently very few oceanic proxy records of sea ice cover with high (less than 10-year) temporal resolution<sup>5,8,9</sup>. Alternatively, terrestrial archives can be used to document the atmospheric forcing of, or atmospheric response to, changes in sea ice cover<sup>5,10,11</sup>. Ice cores preserve signals of atmospheric temperature, moisture source and marine aerosol loadings, all of which may be linked to sea ice conditions<sup>11,12</sup>. Similarly, tree-ring and lake sediments from circum-Arctic sites reflect coastal

summer climate conditions<sup>13</sup>, which may also be linked to sea ice cover<sup>14</sup>. In this study we used a combination of these records to reconstruct variability in Arctic sea ice cover at a five-year resolution since AD 561. We used partial least-square (PLS) regression<sup>15</sup> to reconstruct the summer Arctic sea ice cover, after calibration of the multiproxy network against observational sea ice data over the late nineteenth and twentieth centuries. This approach is founded on the assumption that a component of the temporal covariance of the proxy network is related to the extent of Arctic sea ice cover, and that the relationships between individual proxies and sea ice cover were the same during and before the calibration period. These methods are widely applied for hemispheric-scale palaeoclimate reconstructions, and their merits and shortcomings were recently reviewed<sup>13</sup>.

We used a circum-Arctic array of 69 proxies primarily derived from ice core records ( $n = 52$ ), but also including some long tree-ring chronologies (11), lake sediments (4) and two historical series of sea ice observations (Fig. 1 and Supplementary Table 1). The ice-core proxies are the stable isotope ratio of oxygen ( $\delta^{18}\text{O}$ ), an indicator of air temperature and/or water vapour source and transport history; the percentage of infiltration ice, a proxy for summer warmth, and the concentrations of sea-salt ions ( $\text{Na}^+$  or  $\text{Cl}^-$ ) and oceanic methane-sulphonate, which are linked to sea ice openness and/or sea surface windiness<sup>13</sup>. Tree-ring widths are largely used to infer past summer temperatures, as are lake varve thicknesses<sup>13</sup>. Historical data include the reconstructed ice severity index for Iceland and the mean position of the August ice edge in the Barents Sea. All selected proxies have a



**Figure 1** | Map of the Arctic showing the location and type of proxies used in the reconstruction and ocean sediment cores used for comparison. Red and blue contours respectively delineate the ice edge in August 2007 and 1951, the years of minimum and maximum ice extent from gridded historical sea ice data<sup>3</sup>.

<sup>1</sup>Centro de Estudios Avanzados en Zonas Áridas, Benavente 980, Casilla 554, La Serena, 1720170, Chile. <sup>2</sup>Geological Survey of Canada, NRCAN, Ottawa, Ontario, K1A 0E8, Canada. <sup>3</sup>Norwegian Polar Institute, FRAM – High North Research Centre on Climate and the Environment, NO-9296 Tromsø, Norway. <sup>4</sup>GEOTOP, Université du Québec à Montréal, PO Box 8888, Montréal, Québec, H3C 3P8, Canada. <sup>5</sup>Byrd Polar Research Center, Ohio 43210, Columbus, USA.

time resolution better than five years. Because only terrestrial proxies are considered, coverage is poor in the central Arctic. However, sea ice variability is greatest in the marginal ice zone (the transition zone between the polar ice pack and open ocean), which is well sampled by our proxy network (Fig. 1). Many ice core records are from Greenland. This spatial heterogeneity in proxy coverage is handled by the calibration method used in the reconstruction (see Methods).

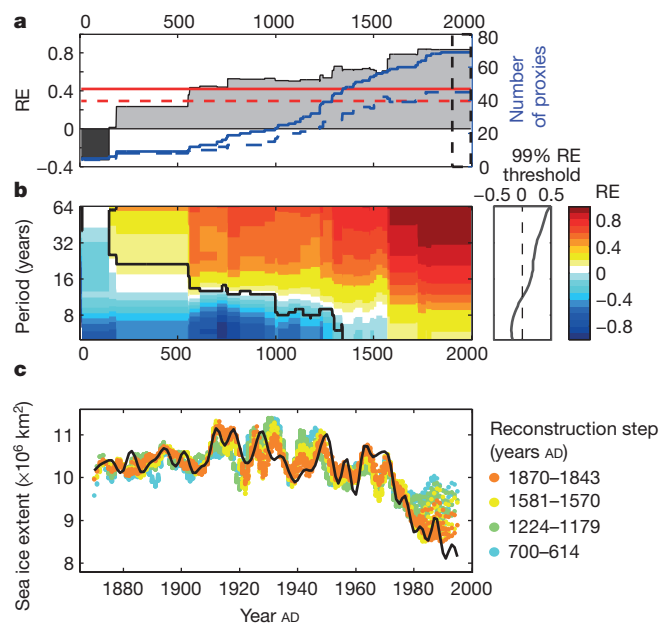
An historical index of late-summer (August) extent of Arctic sea ice (the area of the ocean with at least 15% ice concentration) was used for calibration against the proxy network over the period 1870–1995. The index integrates a gridded compilation of Northern Hemisphere sea ice data<sup>3</sup> with additional data coverage for the Russian Arctic obtained from a separate source<sup>16</sup>. The August ice extent was used because historical data from the Russian Arctic are only available for this month. The total Arctic ice extent in August is close to the September annual minimum, which corresponds approximately to the multi-year ice cover (Supplementary Fig. 1). Additional details on the historical sea ice data sets used are given in the Supplementary Information.

An exploratory Empirical Orthogonal Function (EOF) analysis was first applied on the proxy network over their common time interval, 1843–1995. Unsurprisingly, a dominant temperature-related signal was found. However, a second orthogonal mode was also revealed whose temporal variation tracked the late-summer extent of Arctic sea ice (Supplementary Figs 3 and 4 and Supplementary Information). These findings support our hypothesis that the proxy network contains an embedded climate signal specifically related to sea ice variability, which we sought to extract by using a PLS calibration approach (Methods). A skilful reconstruction of sea ice extent was possible back to AD 561 (Fig. 2a). The Reconstruction of Error skill statistic decreased from 0.84 in the most recent period to 0.43 at

AD 561, after which it became undistinguishable from chance occurrence ( $P < 0.01$ ). Multiscale cross-validation (see Methods) shows significant positive skill at all timescales (period more than five years) after AD 1600, but prior to that, because the number of proxies decreases with increasing age, only the lower frequencies are reconstructed and with decreasing skill. At AD 561 only time periods larger than 25 years are reconstructed, with wider confidence intervals (Fig. 2b, c). A systematic bias towards overestimated reconstructed ice extent is apparent during the past 10–15 years, especially for the earlier proxy-sparse periods (Fig. 2c), which may indicate that the most recent abrupt changes in the Arctic climate system may not be fully captured by the proxies used<sup>17</sup>.

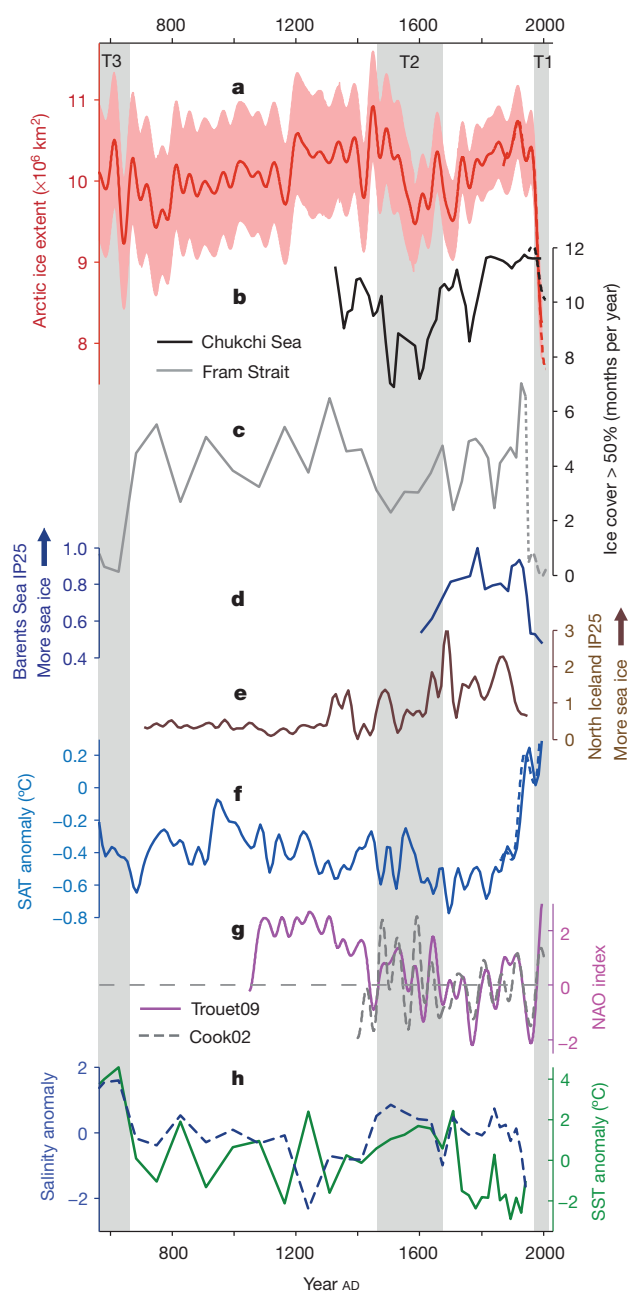
Our proxy-based reconstructed history of late-summer Arctic sea ice extent over the period AD 561–1995 is presented in Fig. 3a along with the observed sea ice record. The reconstruction and observational record were smoothed with a 40-year lowpass filter to highlight the best-resolved frequencies (Fig. 2b). The uncertainty range around the reconstruction widens notably before about AD 1600 as a result of reduced proxy availability and consequent decrease in reconstruction skill. Within this uncertainty range, this reconstruction suggests that the pronounced decline in summer Arctic sea ice cover that began in the late twentieth century is unprecedented in both magnitude and duration when compared with the range of variability of the previous roughly 1,450 years. The most prominent feature is the extremely low ice extent observed since the mid-1990s (T1 in Fig. 3), which is well below the range of natural variability inferred by the reconstruction. Before the industrial period, periods of extensive sea ice cover occurred between AD 1200 and 1450 and between AD 1800 and 1920. Intervals of sustained low extent of sea ice cover occurred before AD 1200, and may be coincident with the so-called Medieval Warm Optimum (roughly AD 800–1300) attested in numerous Northern Hemisphere proxy records<sup>18</sup>, but the pre-industrial minimum occurred before, at about AD 640 (T3 in Fig. 3). Two episodes of markedly reduced sea ice cover also occurred in the late sixteenth and early seventeenth centuries (T2 in Fig. 3). However, by the mid-1990s the observed decrease in sea ice cover had exceeded the lower 95% confidence limit of these pre-historical minima. Our findings support a previous study suggesting that the impact of anthropogenic climate warming on Arctic sea ice became detectable from the early 1990s onwards<sup>19</sup>. The present decline in sea ice is occurring at a pace seen in earlier episodes, but the sustained trend (now nearly 50 years long) is unprecedented in the 1,450-year reconstruction period presented here.

We compared our reconstruction of late-summer sea ice cover derived from terrestrial proxies with independent proxy records of regional sea ice conditions derived from ocean sediment cores. Records for the seasonal duration of sea ice in the Chukchi Sea<sup>9</sup> (Fig. 3b) and Fram Strait<sup>20</sup> (Fig. 3c), reconstructed from dinocyst assemblages, share some of the temporal features of our reconstruction, in particular a pre-industrial minimum around AD 640 in Fram Strait (T3 in Fig. 3), and an apparently marked decrease in sea ice cover between AD 1450 and 1620 in both the Fram Strait and Chukchi Sea (T2 in Fig. 3). Other dinocyst records from northern Baffin Bay and the Beaufort Sea in the Canadian Arctic also show a sea ice minimum around AD 1600 (ref. 21). However, a more detailed comparison is impossible, given the relatively coarse temporal resolution (15–100 years) of oceanic records. The similarity between our Arctic-wide sea ice reconstruction and the dinocyst-based record from Fram Strait suggests that this area, which represents the main connective pathway between the Arctic and the mid latitudes<sup>22</sup>, had a key role in Arctic-wide variations in sea ice extent at interdecadal timescales. Sea ice occurrence in the Barents Sea (Fig. 3d) and in northern Iceland (Fig. 3e) was reconstructed from the sea ice biomarker IP25, which reflects springtime conditions when diatoms bloom<sup>8,23</sup>. The Barents Sea record shows similarities to our Arctic-wide reconstruction and the other records: low values at AD 1600, an increase afterwards and a minimum in the late twentieth century. The record from northern



**Figure 2 | Validation statistics for the reconstruction.** **a**, Shaded grey/black, Reconstruction of Error (RE) statistic; continuous and dashed red lines, 99% and 95% confidence levels on RE from Monte Carlo noise simulations, respectively; blue line, number of potential proxy predictors; dashed blue line, number of proxies retained for prediction; dashed black line, calibration period 1870–1995. **b**, Multiscale validation: the black contour delineates significant RE values ( $P < 0.01$ ); the inset shows the mean RE 99% threshold for significance by period. **c**, Observed ice extent record (black) over the 1870–1995 calibration interval and ice extent predicted by the reduced calibration sets during cross-validation, for four distinct reconstruction steps (coloured points).





**Figure 3 | Comparison between reconstructed late-summer Arctic ice extent and other Arctic sea ice, climate and oceanic proxy records.** **a**, Forty-year smoothed reconstructed late-summer Arctic sea ice extent with 95% confidence interval. **b**, Chukchi Sea ice cover duration reconstructed from core 92-B5 (ref. 9). **c**, Fram Strait sea ice cover duration reconstructed from core JM04 (ref. 20). **d**, Normalized IP25 flux in the BASICC-8 sediment core, a proxy for springtime sea ice occurrence in the western Barents Sea. The timescale used is the average of the two published timescales, corresponding to high and low sedimentation rate alternatives<sup>23</sup>. **e**, Relative abundance of IP25 in core MD99-2275, a proxy for springtime sea ice occurrence along the northern coast of Iceland<sup>8</sup>. **f**, Reconstructed Arctic surface air temperature anomalies<sup>24</sup>. Our sea ice reconstruction includes 9 of the 23 proxies used in this reconstruction. **g**, Two reconstructions of the NAO index<sup>25,26</sup>. The Cook02 record mainly reconstructs interannual variations<sup>26</sup>; its variance was matched to the Trouet09 record, which reflects decadal-scale changes<sup>25</sup>. **h**, Reconstructed sea surface temperature (SST) and salinity in Fram Strait<sup>20</sup>. All records with a time resolution smaller than 40 years were interpolated to 1 year and then smoothed with a 40-year lowpass filter. The dashed lines in **a**, **b**, **c** and **f** are from modern observations. The dotted line in **c** is interpolated between modern observations and the reconstruction. The shaded grey areas T1–T3 delineate the periods of reduced ice extent discussed in the text.

Iceland differs somewhat from the other reconstructions, with a markedly low incidence of sea ice before AD 1200 and a general increase afterwards (Fig. 3e). Accordingly, the Iceland ice severity index is well correlated with the Iceland IP25 record<sup>8</sup>, but not with the Arctic-wide historical ice extent record, and it has a small weight in the reconstruction (Supplementary Table 1). The occurrence of sea ice along the Icelandic coast during winter and spring partly reflects sea ice advection (drift) out of the Arctic rather than ice edge position and sea ice extent.

To evaluate the influence of temperature and dynamic (circulation) forcings on Arctic sea ice extent, we compared our 1,450-year sea ice history with proxy-based reconstructions of Arctic surface atmospheric temperatures (SATs)<sup>24</sup> and of the North Atlantic Oscillation (NAO) index<sup>25,26</sup>. The SAT record shows a sustained cooling trend over about the past 1,250 years followed by an abrupt and pronounced warming in the twentieth century (Fig. 3f). The NAO index also decreased from about AD 1100 onwards, with recent (positive) index values rising sharply in the mid-twentieth century (Fig. 3g). Although the cooling SAT trend is matched by increasing ice extent before AD 1450, both records are otherwise poorly correlated ( $r = -0.12$ ). This reinforces our view that the proxy-based sea ice reconstruction is not just another surrogate Arctic temperature history but instead captures a sea-ice-related climate signal that is not uniquely driven by changes in SATs. Our reconstructed sea ice history is also poorly correlated with the reconstructed NAO indices (Trouet09,  $r = 0.03$ ; Cook02,  $r = -0.13$ ). Hence, the anticorrelation between Arctic sea ice cover and the NAO and related Arctic Oscillation (AO) index reported for recent decades<sup>22</sup> is not readily apparent over the extended sea ice reconstruction. These comparisons suggest that, over the past 1,450 years, changes in the summer extent of Arctic sea ice were not solely forced by SATs or by the NAO/AO, but more probably by a combination of these (and/or other) forcings.

The pronounced decrease in ice cover observed in both our terrestrial and oceanic proxy-based reconstructions between the late fifteenth and early seventeenth centuries occurred during the widespread cooling period known as the Little Ice Age (about AD 1450–1850 (ref. 18)). Reconstructed Arctic SATs show episodes of warming during this period (Fig. 3f), but according to our results the decrease in Arctic sea ice extent during the Little Ice Age was more pronounced than during the earlier Medieval Warm Optimum. A recent climate model simulation of the fifteenth-century Arctic warming (about AD 1470–1520) suggests that it could have been solely driven by enhanced southerly advection of warm air into the Arctic<sup>27</sup>. This is not evident from the reconstructed NAO indices, which alternate between positive and negative and hence do not indicate a markedly increased meridional atmospheric circulation at that time (Fig. 3g). In contrast, positive, synchronous sea surface temperature and salinity anomalies in Fram Strait during this period inferred from dinocyst assemblages (Fig. 3h) suggest increased advection of warm and saline water from the North Atlantic into the Arctic Ocean<sup>20</sup>, and this could have accounted for a decrease in sea ice cover in the Arctic Ocean. Similarly, the pre-industrial minimum in sea ice cover at about AD 640 occurred at the beginning of the Dark Age Cold Period (about AD 600–900 (ref. 28)) under overall cold conditions but was accompanied by increased northward advection of Atlantic water (Fig. 3h). These observations suggest that at multidecadal to centennial timescales, the Arctic sea ice cover during the pre-industrial period may have varied primarily in response to the advection of warm Atlantic water into the Arctic, perhaps as part of internal coupled ocean–atmosphere–sea ice oscillations<sup>29</sup>. Similarly, heat transfer to the Arctic by warm Atlantic water has been shown recently to be unprecedented over the past 2,000 years<sup>28</sup> and may be the main driver for the sustained loss of Arctic sea ice over recent decades. In the present state of knowledge, anthropogenically forced (‘greenhouse gases’) warming stands out as a very plausible cause of the record atmospheric and oceanic warmth of the recent decades, which may soon lead to an ice-free Arctic Ocean in summer.

## METHODS SUMMARY

We used PLS regression, a multivariate calibration method widely used in analytical chemistry<sup>15</sup>, to calibrate the multiproxy ensemble against modern observations of sea ice. The PLS method uses an iterative procedure to generate a set of orthogonal vectors (scores), which are linear combinations of the predictor variables (proxies) and have maximum covariance with the predictand vector (ice extent). PLS constitutes a more flexible approach in our case in which sea ice is not the dominant signal in the proxy network, and it is also ideally suited to deal with a large spatial array of collinear predictor variables, because it permits spatially clustered and/or collinear variables to be combined into fewer vector(s). To account for the uneven number of proxies over time, we used a 'stepwise' calibration approach<sup>17,26</sup>, in which new PLS models were constructed as an increasing number of proxies become available over time. For each model a proxy selection algorithm chose the best subset of predictors based on the available proxies and their contribution to the prediction skill (Methods). At most, 45 proxies were used as predictors in the most recent period, and a minimum of eight proxies were used in the last segment of the reconstruction (Fig. 2a). Model skill was estimated by moving block cross-validation. The Reduction of Error statistic, which predominantly measures the low-frequency skill of a reconstruction<sup>30</sup>, was used as the prediction skill diagnostic as in previous studies<sup>17,26,30</sup>. We also introduced a multiscale validation procedure to assess the predictive skill of each PLS model at different timescales, by applying cross-validation at separate frequency bands (Supplementary Information). The statistical significance of the model prediction skill was determined by Monte Carlo simulations with auto-correlated noise (Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 24 December 2010; accepted 21 September 2011.**

- Serreze, M. C., Holland, M. M. & Stroeve, J. Perspectives on the Arctic's shrinking sea-ice cover. *Science* **315**, 1533–1536 (2007).
- Perovich, D. & Richter-Menge, J. Loss of sea ice in the Arctic. *Annu. Rev. Mar. Sci.* **1**, 417–441 (2009).
- Walsh, J. E. & Chapman, W. L. 20th-century sea-ice variations from observational data. *Ann. Glaciol.* **33**, 444–448 (2001).
- Comiso, J. & Nishio, F. Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *J. Geophys. Res.* **113**, doi:10.1029/2007JC004257 (2008).
- Polyakov, L. *et al.* History of sea ice in the Arctic. *Quat. Sci. Rev.* **29**, 1757–1778 (2010).
- Spielhagen, R. F. *et al.* Enhanced modern heat transfer to the Arctic by warm Atlantic water. *Science* **331**, 450–453 (2011).
- Wang, M. & Overland, J. A sea ice free summer Arctic within 30 years. *Geophys. Res. Lett.* **36**, L07502, doi:10.1029/2009GL037820 (2009).
- Massé, G. *et al.* Abrupt climate changes for Iceland during the last millennium: evidence from high resolution sea ice reconstructions. *Earth Planet. Sci. Lett.* **269**, 565–569 (2008).
- de Vernal, A., Hillaire-Marcel, C., Solignac, S., Radi, T. & Rochon, A. in *Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications* (eds DeWeaver, E. T., Bitz, C. M. & Tremblay, L.-B.) (Geophysical Monograph Series Vol. 80) 27–45 (American Geophysical Union, 2008).
- Macias Fauria, M. *et al.* Unprecedented low twentieth century winter sea ice extent in the Western Nordic Seas since AD 1200. *Clim. Dyn.* **34**, 781–795 (2010).
- Fisher, D. *et al.* Natural variability of Arctic sea ice over the Holocene. *Eos* **87**, 273–275 (2006).
- Grumet, N. S. *et al.* Variability of sea-ice extent in Baffin Bay over the last millennium. *Clim. Change* **49**, 129–145 (2001).
- Jones, P. D. *et al.* High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene* **19**, 3–49 (2009).
- Jacoby, G. C. & Ulan, L. D. Reconstruction of past ice conditions in a Hudson Bay estuary using tree rings. *Nature* **298**, 637–639 (1982).
- Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
- Polyakov, I. *et al.* Long-term ice variability in Arctic marginal seas. *J. Clim.* **16**, 2078–2085 (2003).
- Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).
- Jones, P. & Mann, M. Climate over past millennia. *Rev. Geophys.* **42**, RG2002, doi:10.1029/2003RG000143 (2004).
- Min, S.-K., Zhang, X., Zwiers, F. W. & Agnew, T. Human influence on Arctic sea ice detectable from early 1990s onwards. *Geophys. Res. Lett.* **35**, doi:10.1029/2008GL035725 (2008).
- Bonnet, S., de Vernal, A., Hillaire-Marcel, C., Radi, T. & Husum, K. Variability of sea-surface temperature and sea-ice cover in the Fram Strait over the last two millennia. *Mar. Micropaleontol.* **74**, 59–74 (2010).
- Richerol, T. *et al.* Evolution of paleo sea-surface conditions over the last 600 years in the Mackenzie Trough, Beaufort Sea (Canada). *Mar. Micropaleontol.* **68**, 6–20 (2008).
- Dickson, R. *et al.* The Arctic Ocean response to the North Atlantic Oscillation. *J. Clim.* **13**, 2671–2696 (2000).
- Vare, L. L., Massé, G. & Belt, S. T. A biomarker-based reconstruction of sea ice conditions for the Barents Sea in recent centuries. *Holocene* **20**, 637–643 (2010).
- Kaufman, D. *et al.* Recent warming reverses long-term Arctic cooling. *Science* **325**, 1236–1239 (2009).
- Trouet, V. *et al.* Persistent positive North Atlantic Oscillation mode dominated the medieval climate anomaly. *Science* **324**, 78–80 (2009).
- Cook, E. R., D'Arrigo, R. D. & Mann, M. E. A well-verified, multiproxy reconstruction of the winter North Atlantic Oscillation index since A.D. 1400. *J. Clim.* **15**, 1754–1764 (2002).
- Crespin, E., Goosse, H., Fichefet, T. & Mann, M. The 15th century Arctic warming in coupled model simulations with data assimilation. *Clim. Past* **5**, 389–401 (2009).
- Spielhagen, R. F. *et al.* Enhanced modern heat transfer to the Arctic by warm Atlantic water. *Science* **331**, 450–453 (2011).
- Holland, M. M., Bitz, C. M., Eby, M. & Weaver, A. J. The role of ice-ocean interactions in the variability of the North Atlantic thermohaline circulation. *J. Clim.* **14**, 656–675 (2001).
- Wahl, E. & Ammann, C. Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: examination of criticisms based on the nature and processing of proxy climate evidence. *Clim. Change* **85**, 33–69 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported jointly by the Canadian Foundation for Climate and Atmospheric Sciences (Polar Climate Stability Network) and the Natural Sciences and Engineering Research Council of Canada.

**Author Contributions** C.K. conducted all data preparation and analyses. C.M.Z. and D.A.F. instigated and directed the research and contributed to the interpretation. E.L., A.deV. and L.G.T. contributed data and participated in the interpretation of results. All authors contributed to manuscript preparation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.K. ([christophe.kinnard@cea.ca](mailto:christophe.kinnard@cea.ca)).



## METHODS

**Proxies pre-processing.** The year 1995, beyond which many proxies terminate, was chosen as the end point for the calibration interval. Proxies with missing values in the 1870–1995 calibration interval were filled on the basis of their covariance with the other proxy records, using an EOF-based data-filling procedure<sup>31</sup> (Supplementary Information). All proxies were interpolated to a resolution of 1 year then smoothed to 5-year resolution with a lowpass filter. Extreme proxy values ( $\pm 4$  s.d. around the mean) were deleted and interpolated linearly. Both the proxies and the sea ice series were mean-centred over the calibration interval, and the proxies were also standardized to unit variance over the same interval.

**Calibration methods.** Various methods have been used to calibrate climate observations against a multiproxy network, ranging from simple compositing and variance scaling (CPS), to multivariate climate field reconstructions (CFR)<sup>13,32–34</sup>. In our case, because sea ice cover is not the dominant signal in the proxy network, a simple CPS approach could not be used. PLS regression<sup>15,35</sup> was used to calibrate observations of sea ice extent with the proxy network. A small number of PLS vectors is usually retained to model the predictand and to guard against statistical overfitting. PLS is similar to Canonical Regression<sup>36</sup> and Principal Component Regression (PCR) used in previous palaeoclimate reconstructions<sup>26,36,37</sup>. However, in PLS, vectors are computed and ranked by their covariance with the predictand series, whereas in PCR the vectors are determined solely by the covariance between the predictors, using EOF. The number of potential predictors (proxies)  $k$  decreased back in time, from 69 in AD 1995 to 12 at AD 561 (Fig. 2a). The calibration was performed over the period 1870–1995 ( $n = 126$  years) using a ‘stepwise’ approach<sup>17,26,37</sup>, in which new PLS models were constructed as an increasing number of proxies become available over time. PLS projects the predictor matrix  $X$  (the proxies) into a set of orthogonal score vectors  $t$  that successively maximize the covariance with the dependent variable  $y$  (sea ice extent). A reduced number of scores,  $a < k$ , is retained for the prediction of  $y$ . The PLS model for a single predictand variable is expressed by

$$X = TP' + E$$

$$y = Tq' + f$$

where  $T[n \times a]$  is the score matrix;  $P[k \times a]$  is the loading matrix of  $X$ ;  $E[n \times k]$  is the residual matrix of  $X$ ;  $q$  is the loading matrix of  $y$  and  $f[n \times 1]$  is the vector of  $y$  residuals. The matrix transpose is expressed by the prime symbol. The PLS scores and loadings are calculated successively using the Nonlinear Iterative Partial Least Square (NIPALS) algorithm. For a single predictand variable  $y$ , the NIPALS algorithm is as follows:

- (1) Compute a vector of weights  $w = X'y/(y'y)$ .
- (2) Normalize the weights vector  $w = w/||w||$  (Euclidean norm).
- (3) Calculate the score vector  $t = Xw$ .
- (4) Calculate the loadings for  $X$  and  $y$  as  $P = X't/(t't)$  and  $q = y't/(t't)$ .
- (5) The residuals  $f$  in  $y$  are given by  $f = y - tq'$ .
- (6) The deflated (residual) matrix of predictors is  $E = X - tp'$ .

Each consecutive component (1, ...,  $a$ ), in the model is computed as in step 1 with  $X = E$  and  $y = f$ , and all vectors are stored in matrices. Once the desired number of components has been calculated, the final model can be expressed by

$$\hat{y} = Xb + f$$

$$b = W(P'W)^{-1}q$$

where  $b[k \times 1]$  is the vector of regression coefficients for a PLS model with  $a$  components and  $W[k \times a]$  is the weight matrix of  $X$ .

**Model validation.** Because of its strong persistence and relatively short length ( $n = 126$  years), the observed sea ice record has a low degree of freedom (about 9), which impedes discarding long segments of data for validation, as done elsewhere<sup>17,26</sup>. Instead, we used a moving-block strategy<sup>38</sup>, in which all possible contiguous blocks

of 15 year length were successively removed from the full data set, and PLS was applied on the reduced calibration set and validated on the excluded block. The Reduction of Error (RE) statistic, which predominantly measures the low frequency skill of a reconstruction<sup>30</sup>, was used as the prediction skill diagnostic as in previous studies<sup>17,26,30</sup>. A mean RE statistic was calculated from all excluded blocks. This procedure effectively amounts to successively removing one degree of freedom from the original data before applying calibration. The block size (15 years) was estimated from the autocorrelation function of  $y$  and corresponds to the effective time between two independent observations (Supplementary Information). A backwards variable selection algorithm was incorporated in the cross-validation (CV) procedure to retain the best subset of predictors in each model<sup>39</sup>. A Jack-knife estimate of the variance of the PLS slope vector  $b$  was obtained from CV and the variable with the least significant contribution to the model was removed, before reapplying PLS calibration and CV until all variables retained were found to be significant. At most, 45 proxies were used as predictors in the most recent period, and a minimum of eight proxies were used in the last segment of the reconstruction (Fig. 2a). The 95% confidence interval on the reconstruction was calculated from the cross-validation residuals. We also introduce a multiscale validation procedure to assess the predictive skill of the PLS models at different timescales. The observed ice extent record as well as ice extent predicted by the perturbed PLS models (those trained on the reduced calibration sets during CV) were transformed into separate frequency bands by using the continuous wavelet transform, and a mean RE statistic was calculated at each band (Fig. 2b and Supplementary Information).

**Model dimensionality.** The number of PLS components  $a$  retained for the prediction of  $y$  determines the model dimensionality. Including too many components can lead to statistical over-fitting (modelling noise), whereas too few components will cause under-fitting, meaning that the model does not capture all the variability in the data. A calibration test was first applied to determine model dimensionality: the significance of the covariance between the successive  $y$  and  $t$  vectors was tested by Monte Carlo noise simulation, by randomizing the Fourier phase of  $y$  1,000 times and calculating the product  $t'y$  at each iteration. The number of components  $a$  significantly different from noise ( $P < 0.1$ ) was retained. A validation test was also applied, by examining a plot of the root mean squared error (r.m.s.e.) calculated from cross validation as a function of  $a$ , and choosing the value  $a$  resulting in a clear minimum r.m.s.e. Both methods generally agreed, but when they differed the smallest  $a$  value was retained (Supplementary Fig. 8).

**Model verification.** One-tailed statistical significance tests for the prediction skill of the final PLS models were performed with the same randomization technique described above. A total of 300 surrogate proxy data sets (‘nonsense predictors’) were created and used to predict the record of sea ice extent, and the skill metric (RE) was calculated each time. Models with skill significantly different ( $P < 0.01$ ) from noise were retained (Fig. 2a and Supplementary Fig. 9).

31. Beckers, J. & Rixen, M. EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Ocean. Technol.* **20**, 1839–1856 (2003).
32. Christiansen, B., Schmith, T. & Thejll, P. A surrogate ensemble study of climate reconstruction methods: stochasticity and robustness. *J. Clim.* **22**, 951–976 (2009).
33. Lee, T., Zwiers, F. & Tsao, M. Evaluation of proxy-based millennial reconstruction methods. *Clim. Dyn.* **31**, 263–281 (2008).
34. Mann, M. Climate over the past two millennia. *Annu. Rev. Earth Planet. Sci.* **35**, 111–136 (2007).
35. Naes, T., Isaksson, T., Fearn, T. & Davies, T. *Multivariate Calibration and Classification* (NIR Publications, 2004).
36. Cook, E., Briffa, K. & Jones, P. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *Int. J. Climatol.* **14**, 379–402 (1994).
37. Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* **303**, 1499–1503 (2004).
38. Michaelsen, J. Cross-validation in statistical climate forecast models. *J. Appl. Meteorol.* **26**, 1589–1600 (1987).
39. Martens, H. & Martens, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* **11**, 5–16 (2000).

# Evidence for an oxygen-depleted liquid outer core of the Earth

Haijun Huang<sup>1</sup>, Yingwei Fei<sup>2,3</sup>, Lingcang Cai<sup>4</sup>, Fuqian Jing<sup>1,4</sup>, Xiaojun Hu<sup>1</sup>, Hongsen Xie<sup>5</sup>, Lianmeng Zhang<sup>6</sup> & Zizheng Gong<sup>7</sup>

On the basis of geophysical observations, cosmochemical constraints, and high-pressure experimental data, the Earth's liquid outer core consists of mainly liquid iron alloyed with about ten per cent (by weight) of light elements<sup>1,2</sup>. Although the concentrations of the light elements are small, they nevertheless affect the Earth's core: its rate of cooling, the growth of the inner core, the dynamics of core convection, and the evolution of the geodynamo<sup>3,4</sup>. Several light elements—including sulphur, oxygen, silicon, carbon and hydrogen—have been suggested<sup>2</sup>, but the precise identity of the light elements in the Earth's core is still unclear. Oxygen has been proposed as a major light element in the core on the basis of cosmochemical arguments and chemical reactions during accretion<sup>5,6</sup>. Its presence in the core has direct implications for Earth accretion conditions of oxidation state, pressure and temperature. Here we report new shockwave data in the Fe–S–O system that are directly applicable to the outer core. The data include both density and sound velocity measurements, which we compare with the observed density and velocity profiles of the liquid outer core. The results show that we can rule out oxygen as a major light element in the liquid outer core because adding oxygen into liquid iron would not reproduce simultaneously the observed density and sound velocity profiles of the outer core. An oxygen-depleted core would imply a more reduced environment during early Earth accretion.

To examine the possibility of S and O being the main light elements in the outer core, we conducted shockwave experiments in the Fe–S–O system to produce pressure and density data along the Hugoniot: a sequence of thermodynamic equilibrium states. Both oxygen-rich (6–8 wt% O and 2–3 wt% S)<sup>7</sup> and sulphur-rich (1–3 wt% O and 6–10 wt% S)<sup>8</sup> cores have been proposed, on the basis of geochemical and geophysical arguments, element partitioning or theoretical calculations. We prepared two Fe–S–O starting materials, an oxygen-rich Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> (Fe:O:S = 90:8:2 in weight per cent) and a sulphur-rich Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> (Fe:O:S = 92.5:2.2:5.3 in weight per cent). The Hugoniot data on both samples were obtained using the impedance matching method<sup>9</sup> and are listed in Table 1. A least-squares fit to the

Hugoniot data yields a linear relationship between shock velocity  $U_s$  and particle velocity  $u_p$  whereby  $U_s = 3.71(\pm 0.12) + 1.61(\pm 0.04)u_p$  for Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub>. The Hugoniot data for Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> yielded  $U_s = 3.97(\pm 0.07) + 1.58(\pm 0.03)u_p$  (ref. 10). Figure 1 shows the shock pressure–density relations for Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> and Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub>. The Hugoniot data for Fe, FeO and FeS (refs 11–15) are also shown for comparison.

Using the same Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> starting material and the reverse-impact method<sup>16,17</sup>, we obtained the bulk sound velocities of  $7.15(\pm 0.31)$  km s<sup>−1</sup> and  $7.63(\pm 0.30)$  km s<sup>−1</sup> at 92.6 GPa and 112.9 GPa, respectively. These data represent the sound velocities of solid Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub>. We conducted additional experiments at higher pressures using the optical analyser technique<sup>18</sup> and clearly observed melting at a pressure between 144 GPa and 160 GPa (Supplementary Fig. 6). The bulk sound velocities of liquid Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> were determined up to 208 GPa (Table 2).

We have reported the sound velocities of Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> in our recent study of the melting behaviour of Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub>, determined by the optical analyser technique. Clear melting of Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> was observed between 149 GPa and 167 GPa and the sound velocities at pressures above 167 GPa represent the bulk sound velocities of liquid Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub>.

Figure 2 shows the bulk sound velocities of Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> and Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> as a function of density, compared with the bulk sound velocities of pure iron<sup>19–22</sup>. Adding O and S increases the bulk sound velocity of iron at the same density, but O and S have different effects on the bulk sound velocity. The bulk sound velocities of Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> are higher than those of Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub>, indicating that O has a stronger effect on the bulk sound velocities of iron than S.

To model both the densities and the bulk sound velocities of the liquid outer core, we developed a thermodynamic model that reproduces the observed density  $\rho$  and bulk sound velocity  $C_B$  data. Details of the model are given in the Methods section. Using the equations of state of endmembers (Fe, FeO and FeS), we calculated the densities of Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> and Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> as a function of pressure along the Hugoniot using the additive law<sup>23</sup>. The calculated results agree well with the Hugoniot data for both oxygen-rich and sulphur-rich iron

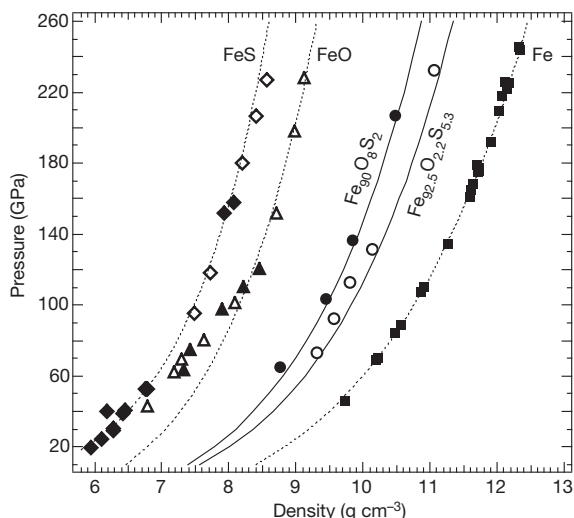
**Table 1 | The Hugoniot data of Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> and Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub>**

Shot number	Impactor	Sample	Impact velocity (km s <sup>−1</sup> )	Initial density (g cm <sup>−3</sup> )	Particle velocity (km s <sup>−1</sup> )	Shock velocity (km s <sup>−1</sup> )	Pressure (GPa)	Density (g cm <sup>−3</sup> )
080903	Cu	Fe <sub>92.5</sub> O <sub>2.2</sub> S <sub>5.3</sub>	3.03 (0.02)	6.88 (0.02)	1.66 (0.02)	6.38 (0.02)	73.0 (0.7)	9.31 (0.09)
081226	Fe <sub>92.5</sub> O <sub>2.2</sub> S <sub>5.3</sub>	LiF	5.48 (0.03)	6.87 (0.02)	1.94 (0.01)	6.92 (0.13)	92.6 (2.3)	9.56 (0.10)
081225	Fe <sub>92.5</sub> O <sub>2.2</sub> S <sub>5.3</sub>	LiF	6.25 (0.02)	6.88 (0.03)	2.21 (0.01)	7.42 (0.12)	112.9 (2.5)	9.80 (0.10)
080911	Cu	Fe <sub>92.5</sub> O <sub>2.2</sub> S <sub>5.3</sub>	4.57 (0.02)	6.87 (0.03)	2.48 (0.02)	7.71 (0.02)	131.6 (1.0)	10.14 (0.09)
080918	Ta	Fe <sub>92.5</sub> O <sub>2.2</sub> S <sub>5.3</sub>	5.77 (0.03)	6.86 (0.03)	3.58 (0.02)	9.45 (0.03)	232.6 (2.4)	11.06 (0.08)
040701*	Cu	Fe <sub>90</sub> O <sub>8</sub> S <sub>2</sub>	2.77 (0.02)	6.70 (0.01)	1.52 (0.02)	6.44 (0.02)	65.4 (0.7)	8.76 (0.08)
040607*	Cu	Fe <sub>90</sub> O <sub>8</sub> S <sub>2</sub>	3.89 (0.02)	6.69 (0.03)	2.12 (0.02)	7.29 (0.06)	103.6 (1.0)	9.44 (0.09)
040611*	Cu	Fe <sub>90</sub> O <sub>8</sub> S <sub>2</sub>	4.69 (0.02)	6.70 (0.02)	2.55 (0.02)	8.00 (0.02)	136.8 (1.0)	9.84 (0.09)
040601*	Ta	Fe <sub>90</sub> O <sub>8</sub> S <sub>2</sub>	5.42 (0.03)	6.69 (0.02)	3.35 (0.03)	9.25 (0.04)	207.4 (2.0)	10.48 (0.07)

\* Data are from ref. 10 and the errors have been re-analysed to be consistent with the current data set.

<sup>1</sup>School of Sciences, Wuhan University of Technology, Wuhan, Hubei 430070, China. <sup>2</sup>Geophysical Laboratory, Carnegie Institution of Washington, Washington, District of Columbia 20015, USA. <sup>3</sup>School of Earth and Space Sciences, Peking University, Beijing 100871, China. <sup>4</sup>National Key Laboratory of Shock Wave and Detonation Physics, Institute of Fluid Physics, China Academy of Engineering Physics, Mianyang, Sichuan 621900, China. <sup>5</sup>Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, Guizhou 550002, China. <sup>6</sup>State Key Lab of Advanced Technology for Materials Synthesis and Processing, Wuhan University of Technology, Wuhan, Hubei 430070, China. <sup>7</sup>National Key Laboratory of Science and Technology on Reliability and Environment Engineering, Beijing Institute of Spacecraft Environment Engineering, Beijing 100094, China.

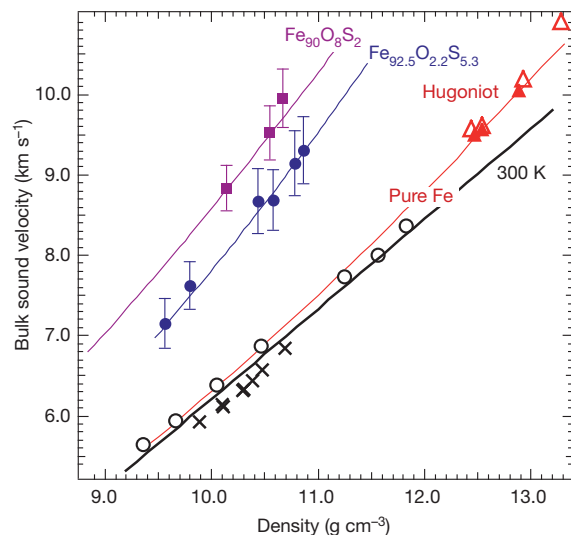




**Figure 1 | The density–pressure relationships for  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_8\text{S}_2$ .** The solid curves represent the calculated Hugoniot curves for  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  (open circles) and  $\text{Fe}_{90}\text{O}_8\text{S}_2$  (solid circles<sup>10</sup>) based on the additive law using the Hugoniot data for Fe, FeS and FeO. Hugoniot data sources are ref. 11 for Fe (solid squares), refs 12 and 13 for FeS (open and solid diamonds, respectively) and refs 14 and 15 for FeO (open and solid triangles, respectively). Dashed lines represent the Hugoniot curves for the high-pressure phases hcp-Fe, B8-FeO and FeS-IV.

alloys in the Fe–S–O system (Fig. 1). Using the same additive law based on the corresponding parameters of Fe, FeO and FeS (listed in table 3 of ref. 18), we calculated the bulk sound velocities from the equations of state and the thermodynamic parameters (see Methods). The calculated bulk sound velocities for  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_8\text{S}_2$  along the Hugoniot reproduce the shockwave measurements well (Fig. 2). We emphasize that the model calculations were based on the equation of state derived from the experimental Hugoniot data and the thermodynamic parameters and are independent of the experimental measurements of the bulk sound velocity. The excellent agreement between the model calculations and experimental measurements provides a self-consistent density and bulk sound velocity data set with which to model the Earth's core.

To compare the measured densities and velocities of  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_8\text{S}_2$  with the observed density and velocity profiles in the outer core, we calculated the densities and bulk sound velocities of  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_8\text{S}_2$  along the adiabatic temperature profile  $T = T_{\text{ICB}}(\rho/\rho_{\text{ICB}})^\gamma$ , using the Grüneisen parameter  $\gamma = 1.5$  and the temperature of the inner core boundary  $T_{\text{ICB}} = 5,400$  K (ref. 18). Figure 3a and b shows the pressure–density and the sound velocity–density relationships, respectively, compared with the PREM model<sup>24</sup>. The density calculations include the density change at melting  $\Delta\rho_{\text{melting}}/\rho_{\text{solid}}$  which is about 0.6% at 330 GPa, using a melting entropy value of  $0.79R$  (where  $R$  is the gas constant)<sup>25</sup>. The value is smaller than the value ( $\sim 1.2\%$ ) for melting of pure iron derived from recent thermodynamic calculations<sup>26</sup>. The densities of  $\text{Fe}_{90}\text{O}_8\text{S}_2$  are significantly lower than the observed values throughout the Earth's outer core, whereas the densities of  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  are slightly higher



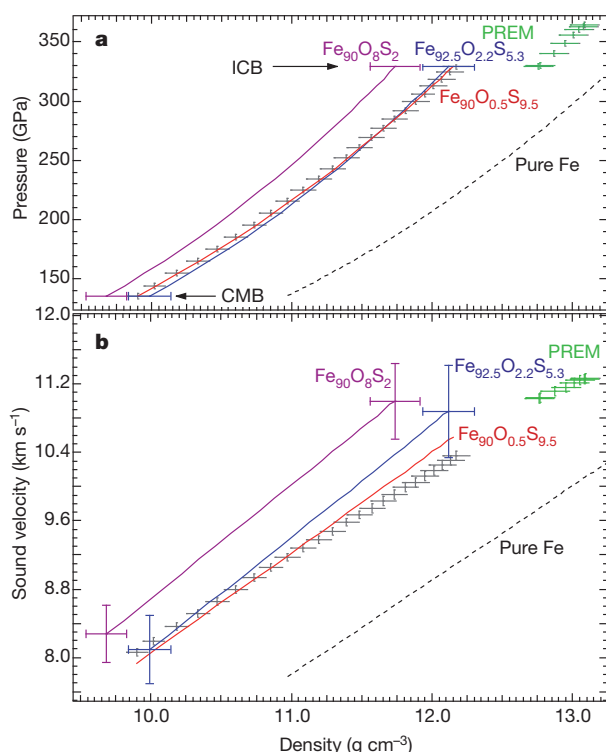
**Figure 2 | The bulk sound velocity as a function of density for  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_8\text{S}_2$ .** The error bars correspond to the experimental uncertainties listed in Table 2. The purple and blue lines represent the calculated bulk sound velocities of  $\text{Fe}_{90}\text{O}_8\text{S}_2$  and  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  along the Hugoniot, respectively, compared with the experimental data for  $\text{Fe}_{90}\text{O}_8\text{S}_2$  (solid squares<sup>18</sup>) and  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  (solid circles). The sound velocities of liquid iron along the Hugoniot (open<sup>19</sup> and solid<sup>20</sup> triangles) are also shown, together with the room-temperature bulk sound velocities of hcp-Fe by static compression (open circles<sup>21</sup> and crosses<sup>22</sup>). The calculated results along the Hugoniot (red line) and at 300 K (black line) are plotted for comparison.

than the observed values at the top of the outer core and cross over near the bottom of the outer core but within the 0.75% density uncertainty of PREM (Fig. 3a). The bulk sound velocities of  $\text{Fe}_{90}\text{O}_8\text{S}_2$  are significantly higher than PREM, deviating from PREM by about 10% at the bottom of the outer core. The bulk sound velocities of  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  match the PREM values at the top of the core, but they deviate from PREM at the bottom of the outer core, about 5% higher than the PREM values, because of a larger  $dC_B/d\rho$  slope (Fig. 3b). Because oxygen has a strong effect on the bulk sound velocity, high oxygen content leads to a bulk sound velocity that is too high compared to PREM for the liquid outer core, especially towards the bottom of the outer core. Therefore, an oxygen-rich liquid outer core would not be consistent with the geophysical observations.

We calculated the density and velocity profiles for a series of compositions with different oxygen and sulphur combinations. The composition of  $\text{Fe}_{90}\text{O}_{0.5}\text{S}_{9.5}$  provides the best fit to both density and velocity profiles of the outer core (Fig. 3). Whereas the density fit is almost perfect, the measured  $dC_B/d\rho$  slope is still slightly higher than observations. We assessed the uncertainties of the calculated bulk sound velocities—4.9% for  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$  and 3.8% for  $\text{Fe}_{90}\text{O}_8\text{S}_2$ —propagated from errors associated with Hugoniot parameters, the Grüneisen parameter, and the specific heat (see Methods). These assessed uncertainties from the model calculations are comparable to the errors on the measured bulk sound velocities, which are  $<5\%$ . We also assessed the uncertainties in density and velocity caused by the temperature uncertainty in the core. A 10% uncertainty in  $T_{\text{ICB}}$

**Table 2 | Sound velocity measurements of  $\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$**

Shot number	Method	Impactor	Sample	Impact velocity (km s <sup>−1</sup> )	Shock velocity (km s <sup>−1</sup> )	Particle velocity (km s <sup>−1</sup> )	Pressure (GPa)	Density (g cm <sup>−3</sup> )	Longitudinal velocity (km s <sup>−1</sup> )	Bulk velocity (km s <sup>−1</sup> )
081226	Reverse-impact	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	LiF	5.48 (0.03)	6.92 (0.13)	3.53 (0.01)	92.6 (2.3)	9.56 (0.10)	8.28 (0.27)	7.15 (0.31)
081225	Reverse-impact	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	LiF	6.25 (0.02)	7.42 (0.12)	4.03 (0.01)	112.9 (2.5)	9.80 (0.10)	8.42 (0.24)	7.63 (0.30)
110426	Optical analyser	Ta	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	4.23 (0.03)	7.94 (0.16)	2.63 (0.02)	143.5 (3.3)	10.27 (0.11)	8.95 (0.41)	Not measured
110510	Optical analyser	Ta	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	4.55 (0.03)	8.25 (0.17)	2.82 (0.02)	159.9 (3.5)	10.44 (0.11)	8.68 (0.40)	8.68 (0.40)
110427	Optical analyser	Ta	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	4.83 (0.03)	8.52 (0.17)	2.99 (0.02)	174.9 (3.8)	10.58 (0.12)	8.69 (0.38)	8.69 (0.38)
110505	Optical analyser	Ta	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	5.25 (0.03)	8.92 (0.18)	3.24 (0.03)	198.5 (4.3)	10.78 (0.13)	9.14 (0.41)	9.14 (0.41)
110511	Optical analyser	Ta	$\text{Fe}_{92.5}\text{O}_{8.3}\text{S}_{5.3}$	5.42 (0.03)	9.08 (0.18)	3.34 (0.03)	208.4 (4.5)	10.86 (0.13)	9.31 (0.42)	9.31 (0.42)



**Figure 3 | Density versus pressure and bulk sound velocity for  $\text{Fe}_{92.5}\text{O}_{2.2}\text{S}_{5.3}$ ,  $\text{Fe}_{90}\text{O}_{8}\text{S}_{2}$ ,  $\text{Fe}_{90}\text{O}_{0.5}\text{S}_{9.5}$  and pure iron along the adiabatic geotherm, compared with the PREM model<sup>24</sup>. a, Density and pressure relationship. b, Density and bulk sound velocity relationship. The error bars indicate 5% and 4% errors in bulk sound velocity for  $\text{Fe}_{92.5}\text{O}_{2.2}\text{S}_{5.3}$  and  $\text{Fe}_{90}\text{O}_{8}\text{S}_{2}$ , respectively, and 1.6% error in density which includes 1.3% error from the uncertainty in the core adiabat and 1% experimental error. The size of the crosses, corresponding to 0.5% uncertainty in sound velocity and 0.75% uncertainty in density, is intended to illustrate the effect of possible uncertainties in PREM on the composition models.**

contributes only a small uncertainty in bulk sound velocity ( $<0.5\%$ ), but a comparable uncertainty in density ( $\sim 1.3\%$ ).

The PREM density of the liquid outer core is about 10% smaller than that of pure iron (Fig. 3a). Adding oxygen and sulphur into iron melt lowers the density. A range of compositions with different amounts of oxygen and sulphur could match the PREM density profile, but the number of combinations that could fit both density and velocity profiles is much smaller. The best fit to both density and velocity profiles allows only 0.5 wt% oxygen in the liquid core for an Fe–S–O core. The maximum amount of oxygen permissible in the liquid outer core is influenced by the estimates of errors on the density and bulk sound velocity. Figure 3 illustrates the importance of error estimates and simultaneous fit to the density and velocity profiles to place bounds on the amount of light elements in the core. For example, the  $\text{Fe}_{92.5}\text{O}_{2.2}\text{S}_{5.3}$  composition gives a reasonable match to the PREM density within the allowed uncertainties (Fig. 3a), but its bulk sound velocities increasingly deviate from PREM velocities towards the bottom of the outer core, reaching a deviation of 5% (Fig. 3b). Our analyses indicate that any Fe–S–O core compositions with more than 2.5 wt% oxygen would not be able to provide good matches to both density and bulk sound velocity profiles of the liquid outer core within the 5% error in bulk sound velocity and 1.6% error in density.

A sulphur-rich core would meet the geophysical requirements. Geochemical constraints usually place 2–3 wt% S in the core<sup>27</sup>. An oxygen-poor core would be consistent with a highly reduced early Earth accretion environment, leaving room for silicon and carbon in the core. Recent continuous accretion models, constrained from siderophile-element-partitioning data, point towards more reduced conditions during early Earth accretion, leading to a core with low oxygen

and high silicon contents<sup>28,29</sup>. This study provides an independent constraint on the oxygen content in the core, placing less than 2.5 wt% oxygen in the liquid outer core with an optimal value of 0.5 wt%. In terms of oxygen content in the core, our result is in a good agreement with a recent core composition model ( $\sim 8$  wt% Si,  $\sim 2$  wt% S and  $\sim 0.5$  wt% O) based on core–mantle element-partitioning data<sup>29</sup>, but it does not support the prediction that oxygen is the major light element in the core on the basis of *ab initio* calculations<sup>7</sup>. The presence of silicon in the core would reinforce the idea of an oxygen-depleted outer core, because this is required by the partitioning data. Replacing some of the sulphur with silicon in the current composition model could easily meet the density requirement of the outer core, but the match to the bulk sound velocity will need to be tested with new shockwave data in the Fe–S–Si system.

## METHODS SUMMARY

The shockwave experiments were performed using a two-stage light gas gun. The Hugoniot data were obtained using the impedance-matching method<sup>3</sup>. The sound velocities were determined using the reverse-impact method<sup>16,17</sup> and the optical analyser technique<sup>18</sup>. The particle velocity history at the impact interface was measured using a displacement interferometer system for any reflector<sup>30</sup>. Using the reverse-impact method, we determined the bulk sound velocity through the relationship between the Lagrangian bulk sound velocity and the particle velocity during unloading. Using the optical analyser technique, we measure the time intervals of the samples with different thicknesses to obtain the catch-up thickness of the sample. The Lagrangian sound velocity was determined from the thickness and velocity of the Ta flyer, and the catch-up thickness and shockwave velocity of the sample.

We developed a thermodynamic model to calculate the densities and bulk sound velocities as a function of pressure and temperature. The thermodynamic parameters for endmembers (Fe, FeO and FeS) were optimised from shockwave data. The calculations include the lattice and electronic contributions to the specific heat and the Grüneisen parameter. Using the additive law, we calculated the densities and bulk sound velocities of alloys in the Fe–S–O system. We also assessed the uncertainty in the calculated bulk sound velocity caused by the Hugoniot parameters, the Grüneisen parameter, the specific heat and core temperature.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 7 December 2010; accepted 3 October 2011.**

- Birch, F. Density and composition of mantle and core. *J. Geophys. Res.* **69**, 4377–4388 (1964).
- Li, J. & Fei, Y. in *Treatise on Geochemistry* (2007 update) (ed. Carlson, R. W.) Vol. 2, 1–31 (Elsevier, 2007).
- Hillgren, V. J., Gessmann, C. K. & Li, J. in *Origin of the Earth and the Moon* (eds Canup, R. M. & Righter, K.) 245–263 (Arizona Univ. Press, 2000).
- Buffett, B. A. Earth's core and the geodynamo. *Science* **288**, 2007–2012 (2000).
- Ringwood, A. E. Composition of the core and implications for origin of the earth. *Geochem. J.* **11**, 111–135 (1977).
- Ohtani, E. & Ringwood, A. E. Composition of the core. I. Solubility of oxygen in molten iron at high temperatures. *Earth Planet. Sci. Lett.* **71**, 85–93 (1984).
- Alfè, D., Gillan, M. J. & Price, G. D. Composition and temperature of the Earth's core constrained by combining *ab initio* calculations and seismic data. *Earth Planet. Sci. Lett.* **195**, 91–98 (2002).
- Helffrich, G. & Kaneshima, S. Seismological constraints on core composition from Fe–O–S liquid immiscibility. *Science* **306**, 2239–2242 (2004).
- Mitchell, A. C. & Nellis, W. J. Shock compression of aluminum, copper, and tantalum. *J. Appl. Phys.* **52**, 3363–3374 (1981).
- Huang, H. J., Jing, F. Q. & Cai, L. C. Studies of the Hugoniot curve for Fe/FeO/FeS mixture. *Chinese J. High Press. Phys.* [in Chinese] **20**, 139–144 (2006).
- Brown, J. M., Fritz, J. N. & Hixson, R. S. Hugoniot data for iron. *J. Appl. Phys.* **88**, 5496–5498 (2000).
- Brown, J. M., Ahrens, T. J. & Shampine, D. L. Hugoniot data for pyrrhotite and the Earth's core. *J. Geophys. Res.* **89**, 6041–6048 (1984).
- Ahrens, T. Equations of state of iron sulfide and constraints on the sulfur content of the Earth. *J. Geophys. Res.* **84**, 985–998 (1979).
- Yagi, T., Fukuoka, K., Takei, H. & Syono, Y. Shock compression of wüstite. *Geophys. Res. Lett.* **15**, 816–819 (1988).
- Jeanloz, R. & Ahrens, T. J. Equation of state of FeO and CaO. *Geophys. J. R. Astron. Soc.* **62**, 505–528 (1980).
- Duffy, T. & Ahrens, T. J. Compressional sound velocity, equation of state, and constitutive response of shock-compressed magnesium oxide. *J. Geophys. Res.* **100**, 529–542 (1995).
- Hu, J., Zhou, X., Tan, H., Li, J. & Dai, C. Successive phase transitions of tin under shock compression. *Appl. Phys. Lett.* **92**, 111905 (2008).



## METHODS

The shockwave experiments were conducted on two Fe–S–O starting materials, with Fe:O:S element ratios of 90:8:2 and 92.5:2.2:5.3 in weight per cent. The starting materials were synthesized from mixtures of high-purity Fe, FeS and Fe<sub>2</sub>O<sub>3</sub> powder at 330 MPa and 850 °C. The Fe<sub>2</sub>O<sub>3</sub> in the synthesized samples was reduced to FeO completely, indicated by Mössbauer spectroscopy and X-ray diffraction measurements. The bulk densities of the samples were measured using the Archimedeian immersion method and listed in Table 1. For the target assembly, we used 12 electrical pins, 60 degrees apart, in two circles around the periphery of the sample on the base plate. The sample was centred on the base plate with an additional 13 electrical pins in two circles and at the centre. All the electrical pins are spring-loaded against a 6-μm Mylar insulating layer. The Hugoniot data were obtained using the impedance-matching method<sup>9</sup>.

The sound velocities of Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> were determined by both the reverse-impact method<sup>16,17</sup> and the optical analyser technique<sup>18</sup>. The reverse-impact method provides measurements of both the longitudinal and bulk sound velocities, whereas the optical analyser technique yields the longitudinal sound velocity before melting and the bulk sound velocity of liquid. The sample Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> (a disk of 30-mm diameter by 2-mm thickness) was used as a flyer to impact directly the LiF single crystal coated with a 3-μm aluminium film, using a two-stage light gas gun. An 8-μm aluminium foil is mounted in the front of the film with epoxy. At impact, shock waves  $D_s$  and  $D_f$  were generated in the sample and the foil, respectively. Supplementary Fig. 1 shows the particle velocity history at the impact interface for two experiments (numbers 081225 and 081226), measured using an all-fibre displacement interferometer system for any reflector (DISAR)<sup>30</sup> with a time resolution of 1 ns. At time  $t_0$ , shock wave  $D_f$  reaches the film–LiF interface. When  $D_s$  reaches the rear surface of the sample, it propagates forward as a rarefaction wave and reaches the foil–window interface at  $t_1$ . The longitudinal sound velocity  $C_L$  is derived by:

$$C_L = \frac{L_s}{(t_1 - t_0) - L_s/D_s} \frac{\rho_0}{\rho} \quad (1)$$

where  $L_s$ ,  $\rho_0$  and  $\rho$  are the thickness of the sample, and the initial and compressed densities, respectively. If we can determine the elastic–plastic transition point at  $t_2$  during unloading, the bulk sound velocities  $C_B$  can also be determined by:

$$C_B = \frac{L_s}{(t_2 - t_0) - L_s/D_s} \frac{\rho_0}{\rho} \quad (2)$$

However, it is often difficult to determine  $t_2$  precisely from the particle velocity history. To obtain the bulk sound velocity precisely, we establish the Lagrangian bulk sound velocity as a function of the particle velocity during unloading, as shown in Supplementary Fig. 2. The intersections of the dashed lines correspond to the Lagrangian bulk sound velocities. The bulk sound velocity is derived by multiplying the Lagrangian bulk sound velocity by  $\rho_0/\rho$ .

Supplementary Fig. 3 shows the experimental configuration for the sound velocity measurements using the optical analyser technique with multiple sample thickness. To measure accurately the time interval of rarefaction sound velocity and shock wave  $D_s$  when they reach the Ta foil, we used the DISAR to monitor the particle velocity of the interface between the foil and LiF window. Supplementary Fig. 4 shows the particle velocity history for experiment 110505. The measured time intervals are 579 ns, 536 ns and 500 ns for samples with the thicknesses of 1.535 mm, 2.336 mm and 3.065 mm, respectively. A linear fit yields the catch-up thickness of the sample,  $d = 12.74$  ( $\pm 0.06$  mm standard deviation) (Supplementary Fig. 5). The Lagrangian sound velocity  $C_s^L$  can be determined by:

$$\frac{d}{D_s^L} = \frac{d_f}{D_f^L} + \frac{d_f}{C_f^L} + \frac{d}{C_s^L} \quad (3)$$

where  $d_f$  is the thickness of the flyer,  $D_f^L$  is the Lagrangian velocity of the flyer and  $D_s^L$  is the Lagrangian shockwave velocity of the sample. The details of the sound velocity determination were discussed in ref. 18.

We also calculated the bulk sound velocity  $C_B$  from a thermodynamic model using:

$$C_B^2 = -V^2 \left( \frac{dP}{dV} \right)_{V,T} + \gamma_{\text{eff}}^2 C_V T \quad (4)$$

The volume  $V$  and  $\left( \frac{dP}{dV} \right)_{V,T}$  are calculated from the equation of state:

$$P(V, T) = P_0(V) + \int_0^{T_m} \frac{\gamma_{\text{eff}}^s}{V} C_V^s dT + \frac{\gamma_{\text{eff}}^l}{V} T_m \Delta S + \int_{T_m}^T \frac{\gamma_{\text{eff}}^l}{V} C_V^l dT \quad (5)$$

where  $\gamma_{\text{eff}}$ ,  $T_m$ ,  $\Delta S$  and  $C_V$  are the effective Grüneisen parameter, the melting temperature, the melting entropy, and the specific heat at constant volume (superscripts s and l indicate solid and liquid), respectively.  $P_0(V)$  is the equation of state at zero Kelvin. In this study, we obtain  $P_0(V)$  from the Hugoniot pressure  $P_H(V, T)$  by:

$$P_0(V) = P_H(V, T) - \int_0^{T_H} \frac{\gamma_{\text{eff}}^s}{V} C_V^s dT \quad (6)$$

The shock temperatures were calculated according to ref. 19. The thermodynamic parameters for endmembers (Fe, FeO and FeS) and for the Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> alloy are listed in ref. 18. For the Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> alloy, we used the following parameters:  $\rho_0 = 6.880$  g cm<sup>−3</sup>,  $C_0 = 3.71 \pm 0.12$  km s<sup>−1</sup>,  $\lambda = 1.61 \pm 0.04$ ,  $\gamma_0 = 1.82 \pm 0.10$ ,  $q = 0.85 \pm 0.10$ ,  $\beta_0 = 0.12 \pm 0.03$  J kg<sup>−1</sup> K<sup>−2</sup>, and  $\kappa = 1.41 \pm 0.08$ . The parameters  $\beta_0$  and  $\kappa$  were used to calculate the electronic contributions to the specific heat<sup>18</sup>.  $C_0$  and  $\lambda$  are the Hugoniot parameters and their uncertainties were derived from fitting the shock-wave and particle-velocity data. The lattice contributions of the Grüneisen parameter is described by  $\gamma_l = \gamma_0(\rho_0/\rho)^q$ .

Using the model described above, we calculated the bulk sound velocity along Hugoniot. The calculations were based on the equation of state derived from the experimental Hugoniot data and the thermodynamic parameters, independently from the bulk sound velocity measurements. The calculated bulk sound velocity for Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> reproduces the measured bulk sound velocity well (Supplementary Fig. 6). The difference in the calculated bulk sound velocity between solid and liquid is negligible (<0.1%).

The errors on the measured bulk sound velocities for Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> range from 3.9–4.6%. The main source of uncertainty in the model calculations is from uncertainty in the derived  $P_0(V)$  caused by Hugoniot parameters, the Grüneisen parameter, and the specific heat. We estimated errors in the calculated bulk sound velocity using:

$$\Delta C_B = \left[ \left( \frac{\partial C_B}{\partial C_0} \Delta C_0 \right)^2 + \left( \frac{\partial C_B}{\partial \lambda} \Delta \lambda \right)^2 + \left( \frac{\partial C_B}{\partial \gamma_0} \Delta \gamma_0 \right)^2 + \left( \frac{\partial C_B}{\partial q} \Delta q \right)^2 + \left( \frac{\partial C_B}{\partial \beta_0} \Delta \beta_0 \right)^2 + \left( \frac{\partial C_B}{\partial \kappa} \Delta \kappa \right)^2 \right]^{1/2} \quad (7)$$

Using the errors associated with these parameters, we calculated the upper and lower bounds of the bulk sound velocities, shown in Supplementary Fig. 6. The propagated errors in sound velocity at outer core pressures are 4.9%, comparable to the largest uncertainty of the measured bulk sound velocity (Supplementary Fig. 6). The measured bulk sound velocities, including their uncertainties, are well represented within the envelope of the calculated bounds on the bulk sound velocities over a wide pressure range, further strengthening the applicability of the model to the core.

The errors on the measured bulk sound velocities for Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> are smaller, ranging from 3.1% to 3.7%. Using the same procedure, we have also calculated the uncertainties in the calculated bulk sound velocity for Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub>. The propagated errors in sound velocity at outer core pressures are 3.8%, smaller than that for the Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub> composition. The uncertainties in the Hugoniot parameters directly contribute to the uncertainties in the derived equation of state at zero Kelvin, having a large influence on the propagated errors in sound velocity. The uncertainties in the Hugoniot parameters for Fe<sub>90</sub>O<sub>8</sub>S<sub>2</sub> are smaller than those for Fe<sub>92.5</sub>O<sub>2.2</sub>S<sub>5.3</sub>, leading to the smaller errors in the calculated bulk sound velocities.

# Changes in plant community composition lag behind climate warming in lowland forests

Romain Bertrand<sup>1,2</sup>, Jonathan Lenoir<sup>3</sup>, Christian Piedallu<sup>1,2</sup>, Gabriela Riofrío-Dillon<sup>1,2</sup>, Patrice de Ruffray<sup>4</sup>, Claude Vidal<sup>5</sup>, Jean-Claude Pierrat<sup>1,2</sup> & Jean-Claude Gégout<sup>1,2</sup>

Climate change is driving latitudinal and altitudinal shifts in species distribution worldwide<sup>1,2</sup>, leading to novel species assemblages<sup>3,4</sup>. Lags between these biotic responses and contemporary climate changes have been reported for plants and animals<sup>5</sup>. Theoretically, the magnitude of these lags should be greatest in lowland areas, where the velocity of climate change is expected to be much greater than that in highland areas<sup>6</sup>. We compared temperature trends to temperatures reconstructed from plant assemblages (observed in 76,634 surveys) over a 44-year period in France (1965–2008). Here we report that forest plant communities had responded to 0.54 °C of the effective increase of 1.07 °C in highland areas (500–2,600 m above sea level), while they had responded to only 0.02 °C of the 1.11 °C warming trend in lowland areas. There was a larger temperature lag (by 3.1 times) between the climate and plant community composition in lowland forests than in highland forests. The explanation of such disparity lies in the following properties of lowland, as compared to highland, forests: the higher proportion of species with greater ability for local persistence as the climate warms<sup>7</sup>, the reduced opportunity for short-distance escapes<sup>8,9</sup>, and the greater habitat fragmentation. Although mountains are currently considered to be among the ecosystems most threatened by climate change (owing to mountaintop extinction), the current inertia of plant communities in lowland forests should also be noted, as it could lead to lowland biotic attrition<sup>10</sup>.

Despite the general pattern of poleward and upward range shifts<sup>1,2</sup>, the timing, magnitude and direction of these biotic responses vary greatly among species and geographic conditions<sup>11–13</sup>. This variation can prevent biotic communities from migrating as intact units in response to climate change<sup>3</sup>. Since the last interglacial, biotic communities have been reshuffled, leading to the formation of novel species assemblages linked to new climatic conditions<sup>3,4</sup>. Recent changes supporting this pattern have occurred within plant and animal communities in response to contemporary climate warming<sup>14–18</sup>. Even though these biotic signals are coherent with climate change<sup>2</sup>, many studies suggest a lag between the magnitude of climate warming and the reshuffling observed in biotic communities<sup>5,12,17–19</sup>. However, comparisons between the extent of the observed biotic signals and that expected based on the intensity of climate warming are still scarce, especially for lowland areas, which offer no short-distance escapes for species facing climate change<sup>8,9</sup>. If biotic communities respond synchronously without lagging behind climate warming, current species compositions would reflect contemporary climatic conditions within both lowland and highland areas. However, the greater velocity of climate change in lowland areas compared to highland areas<sup>6</sup> suggests that reshuffling in biotic communities is more likely to lag behind climate warming in lowland areas compared to highland areas. However, this theory remains unverified.

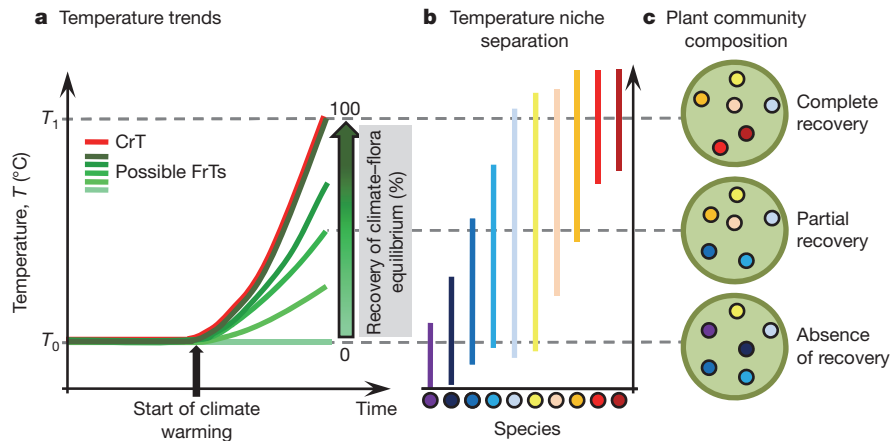
Climate change in France has been characterized by increases in mean annual temperatures of far greater magnitude than the worldwide

average, reaching up to about +1.05 °C on average during 1987–2008 (Supplementary Fig. 2), compared to about +0.35 °C for the worldwide average<sup>20</sup>. Assuming that species assemblages can be used to infer climatic conditions at a given location and time (an approach commonly used by paleoclimatologists<sup>21</sup>, who refer to the relationship as a transfer function), we report a study monitoring plant communities across French forests over 44 years, 1965–2008 (totalling 76,634 surveys; Supplementary Figs 1, 3 and 4). Here we compare temperature trends reconstructed from observed reshuffling of plant assemblages ('floristically reconstructed' temperatures based on 760 species; Supplementary Table 1) with temperature trends reconstructed from instrumental records and climate models ('climatically reconstructed' temperatures based on 237 stations; Supplementary Fig. 5) in lowland (<500 m a.s.l.) and highland forests (500–2,600 m a.s.l.) (Supplementary Methods). Changes in floristically reconstructed temperatures involve reshuffling of plant community composition towards more heat- or cold-demanding species assemblages. Following the current climate warming, small differences between floristically and climatically reconstructed temperatures may indicate important reshuffling of plant community composition, close to the magnitude of climate change (that is, high recovery of the climate–flora equilibrium; Fig. 1). In contrast, large differences between floristically and climatically reconstructed temperatures may indicate that reshuffling in plant communities lags behind climate warming (that is, low recovery of the climate–flora equilibrium; Fig. 1).

Before the contemporary climate warming event (1965–86; Supplementary Fig. 2), we found no significant differences between floristically and climatically reconstructed temperatures in lowland and highland forests (Table 1), suggesting close equilibrium between species composition in biotic assemblages and temperature conditions. In contrast, during the current climate warming period (1987–2008), we found significant differences between floristically and climatically reconstructed temperatures in lowland and highland forests (Table 1), indicating a lag in the response of herbaceous forest species to climate change. In lowland forests, we found no significant increase in floristically reconstructed temperatures between 1965–86 and 1987–2008 (+0.02 °C on average), whereas climatically reconstructed temperatures increased by +1.11 °C on average (Table 1a). In contrast, the floristically reconstructed temperatures of highland forests increased by +0.54 °C, with climate warming reaching +1.07 °C over the same period (Table 1b). This discrepancy indicates an important lag in the response of herbaceous forest species to climate change, being 3.1 times larger in lowland forests (temperature lag = 1.29 °C on average) than in highland forests (temperature lag = 0.42 °C on average; Table 1). In 2008, the most recently investigated year, this lag decreased to an average of 0.28 °C in highland forests but was still 1.26 °C in lowland forests (Fig. 2).

Our results suggest that the recovery of forest plant communities facing climate warming is weak in lowland forests but strong and fast

<sup>1</sup>AgroParisTech, ENGREF, UMR1092 Laboratoire d'Étude des Ressources Forêt-Bois (LERFoB), 14 rue Girardet, F-54000 Nancy, France. <sup>2</sup>INRA, Centre de Nancy, UMR1092 Laboratoire d'Étude des Ressources Forêt-Bois (LERFoB), F-54280 Champenoux, France. <sup>3</sup>Ecoinformatics & Biodiversity Group, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. <sup>4</sup>CNRS, Institut de Biologie Moléculaire des Plantes (IBMP), Université de Strasbourg (UDS), 12 rue du Général Zimmer, F-67084 Strasbourg Cedex, France. <sup>5</sup>Inventaire Forestier National, Château des Barres, F-45290 Nogent-sur-Vernisson, France.



**Figure 1 | Theoretical response of plant communities to climate warming.**

**a**, Floristically (green scale; FrT) and climatically (red; CrT) reconstructed temperature trends over time. The green scale (0–100%) describes different hypothetical floristically reconstructed temperature trends corresponding to increasing intensity of the plant community responses and leading to increasing recovery of the climate–flora equilibrium (measured as the actual change in FrT over time relative to the effective change in CrT over the same time period; see Supplementary Methods for complete formula). **b**, Temperature niche separation among 10 virtual species. The range of temperature requirements for each species is represented by a vertical line. **c**, Three illustrative cases of increasing recovery of the climate–flora equilibrium in plant community composition (based on the 10 virtual species in

**b**). Bottom, absence of climate–flora equilibrium recovery, corresponding to plant communities composed mainly of cold-demanding species, reflecting temperature conditions before climate warming ( $T_0$ ), and leading to an important temperature lag between FrT and CrT. Middle, partial recovery of the climate–flora equilibrium, corresponding to reshuffled plant communities, leading to a mixed assemblage of cold- and heat-demanding species and to reduced temperature lag between FrT and CrT. Top, complete recovery of the climate–flora equilibrium, corresponding to important reshuffling of the plant community, leading to an assemblage of heat-demanding species reflecting the effective temperature increase ( $T_1 - T_0$ ) and to the absence of temperature lag between FrT and CrT. The small circles represent virtual plant species; the three large green disks each depict a community of plants.

in highland forests, reaching  $+0.26^\circ\text{C}$  per decade (Fig. 2). This rate of increase in floristically reconstructed temperatures confirms the results of a previous study, which reported a similar trend of  $+0.25^\circ\text{C}$  per decade for herbaceous forest plants in the French mountains, despite the use of a completely different approach based on the comparison of species' optimum elevation between 1905–85 and 1986–2005<sup>19</sup>. In a Mediterranean forest, the flora turnover observed over the last decade was half that expected from temperature changes<sup>22</sup>, whereas we found that reshuffling of forest plant communities recovered 0% and 60.7% (up to 68.2% in 2008) of the effective temperature increase between 1965–86 and 1987–2008 in lowland and highland areas, respectively. In highland forests, plant communities even completely recovered the effective change in temperature (that is, no significant differences between floristically and climatically reconstructed temperatures) for the years 1991, 1999, 2002, 2004 and 2005 (Fig. 2). Thus, lowland areas seem to be one of the least reactive terrestrial ecosystems to climate warming with respect to forest plant communities.

Several possible factors may explain the lower recovery of forest plant communities in response to climate warming in lowland versus highland areas. First, the larger proportion of cosmopolitan and thermophilous species (Supplementary Methods) in lowland communities (75.2% and 15.3% on average) compared to highland communities (67.2% and 8.8% on average) may result in a greater tolerance of lowland communities to climate warming throughout local persistence<sup>7</sup>. We did not find any significant changes in the composition of lowland

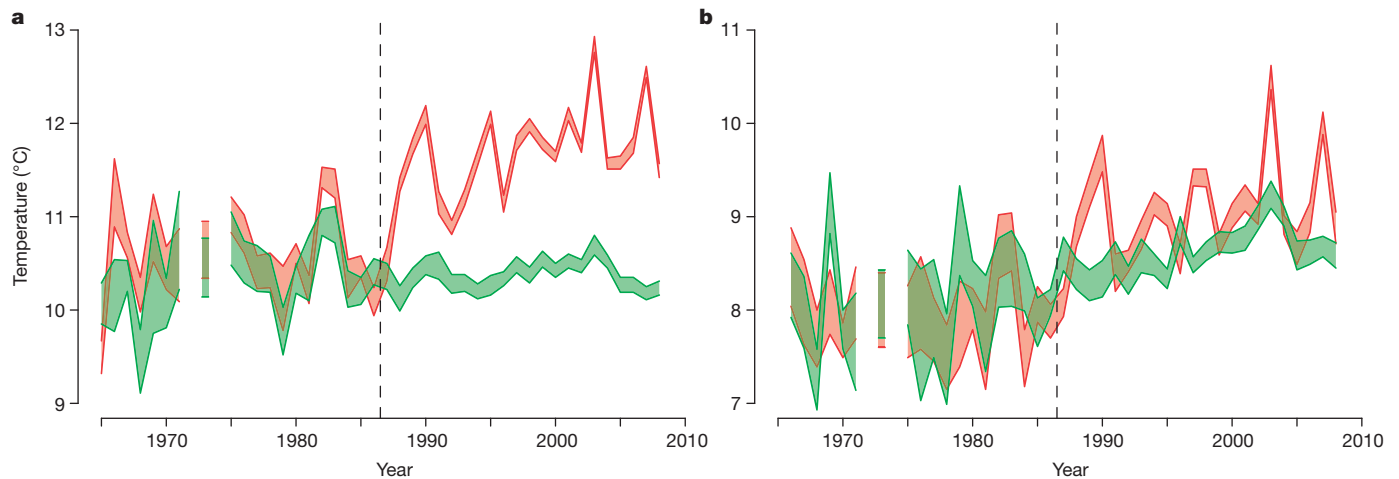
plant communities between 1965–86 and 1987–2008 (Fig. 3), whereas the proportion of mountainous and alpine species in highland plant communities decreased significantly ( $-4.6\%$  on average), to the benefit of cosmopolitan species ( $+5.1\%$  on average). In addition, we suggest that compensatory changes in demographic rates may buffer population dynamics against the negative effects of climate warming, causing a temporary delay in species turnover, especially in lowland plant communities (for example, a higher growth of individuals can compensate for the lower survival and recruitment rates, allowing persistence of threatened populations)<sup>12</sup>. Second, highland forests offer shorter-distance escapes for species facing climate warming compared to lowland forests<sup>8,9</sup>, especially for forest plants that have strongly limited dispersal abilities<sup>23</sup>. Thus, species distributions are expected to migrate  $+1.1\text{ km}$  (median value; 5th to 95th percentile range ( $\text{PR}_{5,95}$ ) =  $[0.4; 15]$ ) upward in French highland forests, but  $+35.6\text{ km}$  (median value;  $\text{PR}_{5,95}$  =  $[2.1; 137.6]$ ) northward in French lowland forests to track their climatic niches according to the contemporary climate warming event (1987–2008; Supplementary Fig. 9). Third, the migration rate of plant species may be delayed by the level of habitat fragmentation<sup>8,23</sup>. Forest patches exceeding  $5\text{ km}^2$  cover a larger proportion of highland areas (67.3%) than lowland areas (29.1%; Supplementary Fig. 10a). Moreover, forest patches are closer in highland areas (median proximity index (PI) =  $3,813.7$ ;  $\text{PR}_{5,95}$  =  $[15.3; 122,034.3]$ ; see Supplementary Methods for details) than in lowland areas (median PI =  $31.6$ ;  $\text{PR}_{5,95}$  =  $[1.3; 14,644.2]$ ; Supplementary Fig. 10b). All of these factors are not mutually exclusive and may,

**Table 1 | Comparison of floristically and climatically reconstructed temperatures.**

a Lowland				b Highland			
Reconstruction	Period			Reconstruction	Period		
	1965–1986 ( $P_1$ )	1987–2008 ( $P_2$ )	$P_2 - P_1$		1965–1986 ( $P_1$ )	1987–2008 ( $P_2$ )	$P_2 - P_1$
CrT	10.54 (1.2)	11.66 (1.23)	1.11 (0.02)**	CrT	7.98 (1.8)	9.05 (1.8)	1.07 (0.03)**
FrT	10.35 (1.23)	10.37 (1.22)	0.02 (0.02)	FrT	8.09 (1.77)	8.63 (1.61)	0.54 (0.03)**
FrT – CrT	–0.19 (0.87)	1.29 (0.99)**		FrT – CrT	0.11 (1.12)	0.42 (1.23)*	

Shown are mean temperatures with standard deviations (in brackets) computed from the 1,000 floristically (FrT) and climatically (CrT) reconstructed temperature trends.  $P_1$ , Value for 1965–86;  $P_2$ , value for 1987–2008. **a**, Comparison for French lowland forests ( $<500\text{ m a.s.l.}$ ). **b**, Comparison for French highland forests ( $500\text{–}2,600\text{ m a.s.l.}$ ). Significance of temperature differences between periods (Student's  $t$  test) and between modelling approaches (Student's paired  $t$  test) are indicated: \* $P < 0.01$ , \*\* $P < 0.001$  for more than 95% of the 1,000 reconstructed temperatures.





**Figure 2 | Comparison of floristically (green) and climatically (red) reconstructed temperature trends between 1965 and 2008. a,** Trends in lowland forest plant communities (<500 m a.s.l.). **b,** Trends in highland forest plant communities (500–2,600 m a.s.l.). The thickness of lines shows the range

of reconstructed temperature trends ( $n = 1,000$  trends). Dashed lines indicate the start of the contemporary climate warming period (1987–2008). Breaks in trends are due to no sample convergence for the years 1965 (in highland areas), 1972 and 1974 (in both lowland and highland areas).

independently or combined, explain our results showing that reshuffling of plant communities in lowland forests strongly lag behind climate warming.

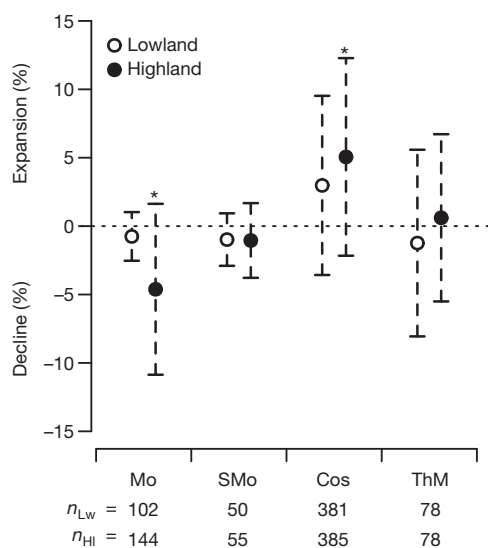
In this study, we linked reshuffling observed in plant assemblages in French forests to temperature changes between 1965 and 2008. However, water stress may also drive species range shifts<sup>11,22</sup>. In California, the core of the altitudinal range of 46 out of 64 forest plant species shifted downward in response to regional changes in climatic water balance between 1935 and 2005<sup>11</sup>. In contrast, no significant changes in the level of annual precipitation was observed throughout the French territory between 1965–86 and 1987–2008 (mean value of

precipitation changes [s.d.] = +1.04 mm [48.56] in lowland areas,  $P = 0.35$ , Student's  $t$  test for paired samples;  $-0.93$  mm [60.09] in highland areas,  $P = 0.44$ , Student's  $t$  test for paired samples; Supplementary Fig. 2). Thus, the absence of changes in precipitation patterns over the 44-year study period is unlikely to explain the differences of reshuffling between lowland and highland forest plant communities. Similarly, acclimation or adaptation processes may also explain the differences we observed between floristically and climatically reconstructed temperatures. The large difference in temperature recovery between highland and lowland forest plant communities may indicate distinct interchanges between migration and adaptation, with migration being more prevalent in highland forests and adaptation more prevalent in lowland forests. Nevertheless, such processes are probably insufficient to mitigate completely the effects of the high temperature increase in just 22 years, especially at the scale of a large pool of perennial species in a fragmented landscape<sup>24,25</sup>.

Although a potential lag between the magnitude of climate warming and the extent of changes within biotic communities has already been suggested<sup>5,12,17–19</sup>, we provide the first (to our knowledge) assessment of such a lag and show that it is more important in lowland areas. This finding has important implications in terms of conservation issues. To date, a strong focus has been placed on mountaintop extinctions for species facing climate warming<sup>10</sup>. However, mountaintop extinctions are just one example, although important, of the likely consequences of contemporary species range shifts in highland areas. Highland areas offer short-distance escapes for species facing climate change<sup>8,9</sup> and provide ideal refuges for plant species to track their climatic requirements<sup>10</sup>, which could lead to the formation of novel species assemblages. In lowland areas, the current inertia of forest plant communities suggests delayed extinction and immigration in the face of climate warming<sup>26</sup> due to species tolerance and low migration rate. Considering the velocity of predicted future climate change<sup>6</sup>, the low dispersal abilities of forest plants<sup>23</sup> and the low opportunity for short-distance escape to track climate changes in lowland areas<sup>8,9</sup>, the current inertia of forest plant communities will not be indefinite. Consequently, abrupt reshuffling of these plant communities may occur in the future, possibly leading to lowland biotic attrition<sup>10</sup> with important changes expected in ecosystem functioning<sup>27</sup>. Therefore, we recommend that research efforts be focused on understanding the implications of climate change in lowland areas.

## METHODS SUMMARY

To achieve our aim, we compiled a complete species list including the year and location of 79,621 floristic surveys (44,509 and 35,112 in lowland and highland



**Figure 3 | Compositional changes in the plant communities of lowland and highland forests according to four different biogeographic groups.** Mean shifts in the proportions of plant communities are shown (data points) with standard deviations (error bars) estimated from 1,000 floristic samples used to reconstruct temperatures. The significance of changes from the null hypothesis of zero shift is displayed (\* $P < 0.05$  for more than 95% of the floristic samples; Wilcoxon paired signed-rank test). The number of species analysed in lowland ( $n_{Lw}$ ) and highland ( $n_{Hl}$ ) plant communities are displayed below the figure. Mo, mountainous to alpine species; SMO, sub-mountainous to lowland species; Cos, cosmopolitan species; ThM, thermophilous to Mediterranean species. See Supplementary Methods and Supplementary Table 1 for more details about the different biogeographic groups.

areas, respectively) that were recorded across the French forest territory between 1965 and 2008. We computed a 1-km<sup>2</sup> grid of the mean annual temperature across the French territory for each year between 1965 and 2008 using a modelling approach based on 237 non-interrupted meteorological stations ( $R^2 = 0.93$  for 13,620 independent temperature observations; Supplementary Figs 5 and 6)<sup>28</sup>. These climatically reconstructed temperatures were extracted at the location and year of each of the 79,621 floristic surveys. We then subsampled a training data set of 2,987 floristic surveys before the recent climate warming event (1975–85) to calibrate a transfer function using a combination of weighted averaging partial least squares<sup>29</sup> and Breiman's random forest<sup>30</sup> to infer temperatures from the plant assemblages ( $R^2 = 0.83$  in a validation data set; Supplementary Fig. 7). These floristically reconstructed temperatures were then predicted across the French territory between 1965 and 2008 using the 76,634 floristic surveys. This predictive data set was bootstrapped separately in lowland and highland areas for each year to control for artificial temperature variations associated with inter-annual variations in floristic survey locations (Supplementary Fig. 8). We reconstructed 1,000 pairs of floristically and climatically reconstructed mean annual temperature trends separately for lowland and highland areas. We tested temperature differences in each pair of trends between 1965–86 and 1987–2008, as well as annually. To discuss our results, we assessed changes in plant community composition in four biogeographic groups of species. We then computed the mean distance separating equal isotherms between periods, in addition to performing independent spatial analysis of the degree of habitat fragmentation in lowland and highland forest areas (see Supplementary Methods for more details).

Received 25 May; accepted 9 September 2011.

Published online 19 October 2011.

1. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
2. Rosenzweig, C. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L. *et al.*) 79–131 (Cambridge Univ. Press, 2007).
3. Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ* **5**, 475–482 (2007).
4. Wing, S. L. *et al.* Transient floral change and rapid global warming at the Paleocene-Eocene boundary. *Science* **310**, 993–996 (2005).
5. Davis, M. B. in *Community Ecology* (eds Diamond, J. & Case, T. J.) 269–284 (Harper and Row, 1986).
6. Loarie, S. R. *et al.* The velocity of climate change. *Nature* **462**, 1052–1055 (2009).
7. Thuiller, W., Lavorel, S., Araujo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proc. Natl Acad. Sci. USA* **102**, 8245–8250 (2005).
8. Jump, A. S., Matyas, C. & Penuelas, J. The altitude-for-latitude disparity in the range retractions of woody species. *Trends Ecol. Evol.* **24**, 694–701 (2009).
9. Scherrer, D. & Korner, C. Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *J. Biogeogr.* **38**, 406–416 (2011).
10. Colwell, R. K., Brehm, G., Cardelus, C. L., Gilman, A. C. & Longino, J. T. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322**, 258–261 (2008).
11. Crimmins, S. M., Dobrowski, S. Z., Greenberg, J. A., Abatzoglou, J. T. & Mynsberge, A. R. Changes in climatic water balance drive downhill shifts in plant species' optimum elevations. *Science* **331**, 324–327 (2011).
12. Doak, D. F. & Morris, W. F. Demographic compensation and tipping points in climate-induced range shifts. *Nature* **467**, 959–962 (2010).
13. Lenoir, J. *et al.* Going against the flow: potential mechanisms for unexpected downslope range shifts in a warming climate. *Ecography* **33**, 295–303 (2010).
14. Cannone, N., Sgorbati, S. & Guglielmin, M. Unexpected impacts of climate change on alpine vegetation. *Front. Ecol. Environ* **5**, 360–364 (2007).
15. Hillebrand, H., Soininen, J. & Snoeijs, P. Warming leads to higher species turnover in a coastal ecosystem. *Glob. Change Biol.* **16**, 1181–1193 (2010).
16. le Roux, P. C. & McGeoch, M. A. Rapid range expansion and community reorganization in response to warming. *Glob. Change Biol.* **14**, 2950–2962 (2008).
17. Lenoir, J., Gegout, J.-C., Dupouey, J.-L., Bert, D. & Svenning, J.-C. Forest plant community changes during 1989–2007 in response to climate warming in the Jura Mountains (France and Switzerland). *J. Veg. Sci.* **21**, 949–964 (2010).
18. Moritz, C. *et al.* Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science* **322**, 261–264 (2008).
19. Lenoir, J., Gegout, J.-C., Marquet, P. A., de Ruffray, P. & Brisse, H. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**, 1768–1771 (2008).
20. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J. Geophys. Res. Atmos.* **111**, D12106, 1–21 (2006).
21. Heikkilä, M. & Seppä, H. A 11,000 yr palaeotemperature reconstruction from the southern boreal zone in Finland. *Quat. Sci. Rev.* **22**, 541–554 (2003).
22. Vennetier, M. & Ripert, C. Forest flora turnover with climate change in the Mediterranean region: a case study in Southeastern France. *For. Ecol. Mgmt* **258**, S56–S63 (2009).
23. Primack, R. B. & Miao, S. L. Dispersal can limit local plant-distribution. *Conserv. Biol.* **6**, 513–519 (1992).
24. Huntley, B. Evolutionary response to climatic change? *Heredity* **98**, 247–248 (2007).
25. Jump, A. S. & Penuelas, J. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol. Lett.* **8**, 1010–1020 (2005).
26. Jackson, S. T. & Sax, D. F. Balancing biodiversity in a changing environment: extinction debt, immigration credit and species turnover. *Trends Ecol. Evol.* **25**, 153–160 (2010).
27. Walther, G. R. Community and ecosystem responses to recent climate change. *Phil. Trans. R. Soc. B* **365**, 2019–2024 (2010).
28. Ninyerola, M., Pons, X. & Roure, J. M. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Climatol.* **20**, 1823–1841 (2000).
29. ter Braak, C. J. F. & van Dam, H. Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* **178**, 209–223 (1989).
30. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J.-D. Bontemps and P. Mérian for comments; V. Pérez, F. Lebourgeois and E. K. Cavalheri for help in the compilation of the meteorological database; V. Pérez for technical support in GIS; I. Seynave for management of the EcoPlant database; H. Brisse, J. Drapier and F. Morneau for contributions to the Sophy and NFI databases; and all who have participated in the conception of the EcoPlant, Sophy and NFI databases. The phytoecological database (EcoPlant) was funded by the French Institute of Agricultural, Forest and Environmental Engineering (ENGREF, AgroParisTech), the National Forest Department (ONF), and the French Agency for Environment and Energy Management (ADEME). J.L. acknowledges a grant from the Danish Council for Independent Research – Natural Sciences (272-07-0242 to J.-C. Svenning). This study was funded through a PhD grant to R.B. by ADEME and the Regional Council of Lorraine.

**Author Contributions** R.B. designed the study, methodology and modelling approach, performed all the statistical analysis and wrote the paper; P.d.R. provided the Sophy database; C.V. provided the NFI database; J.-C.G. provided the EcoPlant database, helped to design the methodology and supervised the work; R.B. and G.R. contributed equally to format the floristic database; J.-C.P. advised the use of the Breiman's random forest regression to infer temperatures from the plant assemblages; R.B. and C.P. contributed equally to compute the climate model of historic temperature prediction; J.L. contributed actively to improve the clarity of the paper. All authors discussed and commented on the results.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.B. ([romain.bertrand@engref.agroparistech.fr](mailto:romain.bertrand@engref.agroparistech.fr)).

# The earliest evidence for anatomically modern humans in northwestern Europe

Tom Higham<sup>1</sup>, Tim Compton<sup>2</sup>, Chris Stringer<sup>2</sup>, Roger Jacobi<sup>2,3,†</sup>, Beth Shapiro<sup>4</sup>, Erik Trinkaus<sup>5</sup>, Barry Chandler<sup>6</sup>, Flora Gröning<sup>7</sup>, Chris Collins<sup>2</sup>, Simon Hillson<sup>8</sup>, Paul O'Higgins<sup>9</sup>, Charles FitzGerald<sup>10</sup> & Michael Fagan<sup>7</sup>

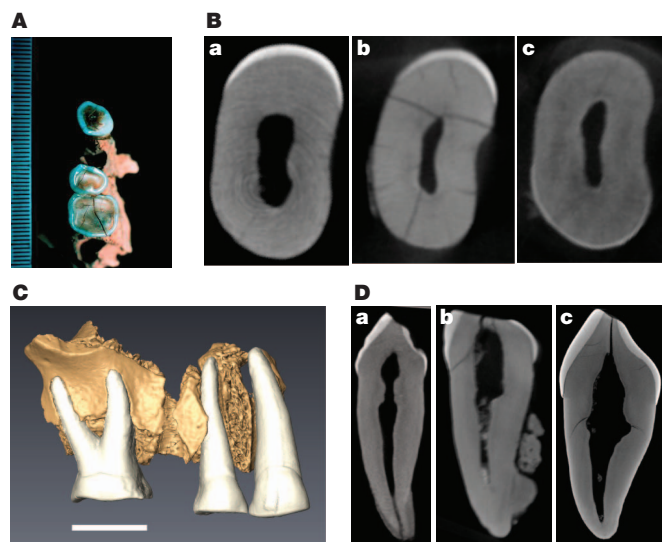
The earliest anatomically modern humans in Europe are thought to have appeared around 43,000–42,000 calendar years before present (43–42 kyr cal BP), by association with Aurignacian sites and lithic assemblages assumed to have been made by modern humans rather than by Neanderthals. However, the actual physical evidence for modern humans is extremely rare, and direct dates reach no farther back than about 41–39 kyr cal BP, leaving a gap. Here we show, using stratigraphic, chronological and archaeological data, that a fragment of human maxilla from the Kent's Cavern site, UK, dates to the earlier period. The maxilla (KC4), which was excavated in 1927, was initially diagnosed as Upper Palaeolithic modern human<sup>1</sup>. In 1989, it was directly radiocarbon dated by accelerator mass spectrometry to 36.4–34.7 kyr cal BP<sup>2</sup>. Using a Bayesian analysis of new ultrafiltered bone collagen dates in an ordered stratigraphic sequence at the site, we show that this date is a considerable underestimate. Instead, KC4 dates to 44.2–41.5 kyr cal BP. This makes it older than any other equivalently dated modern human specimen and directly contemporary with the latest European Neanderthals, thus making its taxonomic attribution crucial. We also show that in 13 dental traits KC4 possesses modern human rather than Neanderthal characteristics; three other traits show Neanderthal affinities and a further seven are ambiguous. KC4 therefore represents the oldest known anatomically modern human fossil in northwestern Europe, fills a key gap between the earliest dated Aurignacian remains and the earliest human skeletal remains, and demonstrates the wide and rapid dispersal of early modern humans across Europe more than 40 kyr ago.

The dispersal of the first anatomically modern humans (AMH) into Europe has long been associated with the transition from the Middle to the Upper Palaeolithic and linked with the subsequent disappearance of Neanderthals. Numerous sites attest to the change from late Neanderthal Mousterian or transitional industries to Early or proto-Aurignacian technocomplexes, evidenced by the presence of blade-based lithic assemblages and a wide range of bone implements and ornaments. However, the skeletal evidence linking AMH with this earliest phase is extremely fragmentary in comparison with remains from the later Evolved Aurignacian period. The earliest direct date on an AMH bone (42–37.8 kyr cal BP) comes from Peștera cu Oase, Romania, but the date ranges are wide and the human fossils are not associated with any tool industry<sup>3</sup>.

The site of Kent's Cavern, Torquay, UK, is of great importance because it contains an early, directly dated human maxilla (KC4; Fig. 1A) from a site at the maximal north-westerly range of the European Aurignacian. Stratified sediments ranging from Middle Pleistocene to Holocene in age were excavated in the site, within a

cave earth sealed periodically by stalagmite deposits<sup>4</sup>. The maxilla was found in 1927 at a depth of 10 ft 6 inch (3.23 m) beneath a key 'granular stalagmite' used as a datum during excavations undertaken between 1926 and 1941 by the Torquay Natural History Society<sup>5</sup>. Below it were found two blades similar to those discovered in Aurignacian industries, and deeper still were found two blades that resemble those from Initial Upper Palaeolithic industries of the Lincombian–Ranisian–Jerzmanowician complex, which are tentatively associated with Neanderthals<sup>6,7</sup>. Keith, who analysed the maxilla, considered it to be of anatomically modern type<sup>1</sup>, and interest in the fossil further increased in 1989 when direct dating by accelerator mass spectrometry (AMS) confirmed its early Upper Palaeolithic age (30,900 ± 900 radiocarbon years (yr <sup>14</sup>C) BP; ref. 2). The age fitted then-current views on the dating and modern human associations of the Aurignacian industry in north-western Europe.

However, recent research into the chronology of the Middle to Upper Palaeolithic has suggested that there are severe problems with the accuracy of many <sup>14</sup>C dates from this period, especially those



**Figure 1 | Morphology of the KC4 maxilla.** A, Original dentition within the KC4 maxilla as Keith<sup>1</sup> would have seen it. B, Third premolar (P<sup>3</sup>) pulp chamber shape of Neanderthal (a), KC4 (b) and modern human (c). C, CT-based three-dimensional model of the KC4 maxilla showing the new reconstruction of the specimen (scale bar, 1 cm). D, Upper canine pulp chamber shapes of Neanderthal (a), KC4 (b) and modern human (c). Neanderthal specimens are from the site at Krapina, Croatia; modern humans are from Gough's Cave, UK. See Supplementary Methods for further details.

<sup>1</sup>Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QY, UK. <sup>2</sup>The Natural History Museum, Cromwell Road, London SW7 5BD, UK. <sup>3</sup>Department of Prehistory and Europe (Quaternary Section), The British Museum, London N1 5QJ, UK. <sup>4</sup>Department of Biology, The Pennsylvania State University, 208 Mueller Laboratory, University Park, Pennsylvania 16802, USA. <sup>5</sup>Department of Anthropology, Campus Box 1114, Washington University, St Louis, Missouri 63130, USA. <sup>6</sup>Torquay Museum, 529 Babbacombe Road, Torquay TQ1 1HG, UK. <sup>7</sup>The Department of Engineering, University of Hull, Kingston upon Hull HU6 7RX, UK. <sup>8</sup>Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK. <sup>9</sup>Centre for Anatomical and Human Sciences, Hull York Medical School, The University of York, Heslington, York YO10 5DD, UK. <sup>10</sup>Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L9, Canada.

†Deceased.

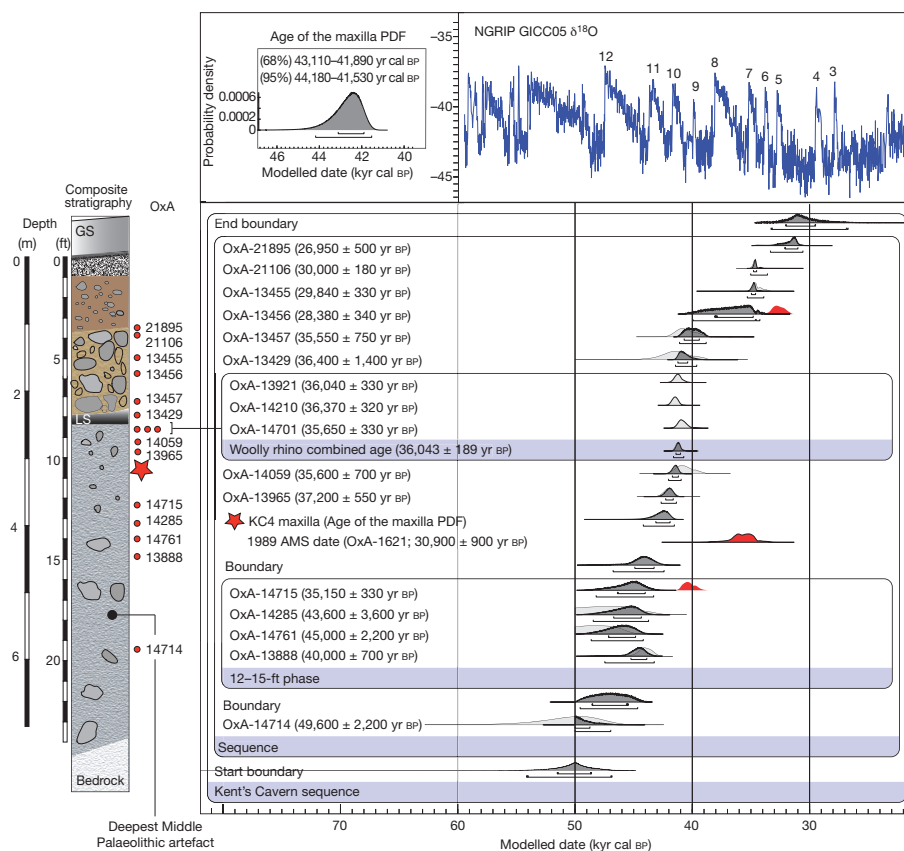


produced early in the developmental stages of the technique<sup>8,9</sup>. Contamination problems with many of the existing AMS determinations on bone dating from the Middle and Upper Palaeolithic before 30 kyr <sup>14</sup>C BP can be reduced considerably by the use of ultrafiltration during collagen extraction<sup>9,10</sup>. We used this technique to reinvestigate the age of KC4, first by dating faunal material above and below the maxilla find spot, and then by attempting a direct re-dating of a tooth from the maxilla itself.

Faunal remains from the Torquay Natural History Society excavations that could be matched with the descriptions and contexts given in the *Kent's Cavern Journal* (1926–1932) were selected for radiocarbon dating. AMS dates for fauna from two of the excavation trenches (trenches B and C, eventually linked together by the excavators) are shown in stratigraphic order in increasing depth below the granular stalagmite in Fig. 2 (see also Supplementary Table 1). The results show that the initial AMS determination for the maxilla is too young, probably because of the presence of trace animal glues not effectively removed during chemical pretreatment<sup>8</sup> (Supplementary Methods). To constrain more precisely the age of the maxilla, we used the new dates and independent stratigraphic depth data to construct a Bayesian model using OxCal 4.1 software<sup>11</sup> and the INTCAL09 <sup>14</sup>C calibration curve<sup>12</sup>. This approach allows the relative stratigraphic information from the site to be modelled mathematically along with the calibrated radiocarbon likelihoods (Fig. 2). We are confident that dates corresponding to

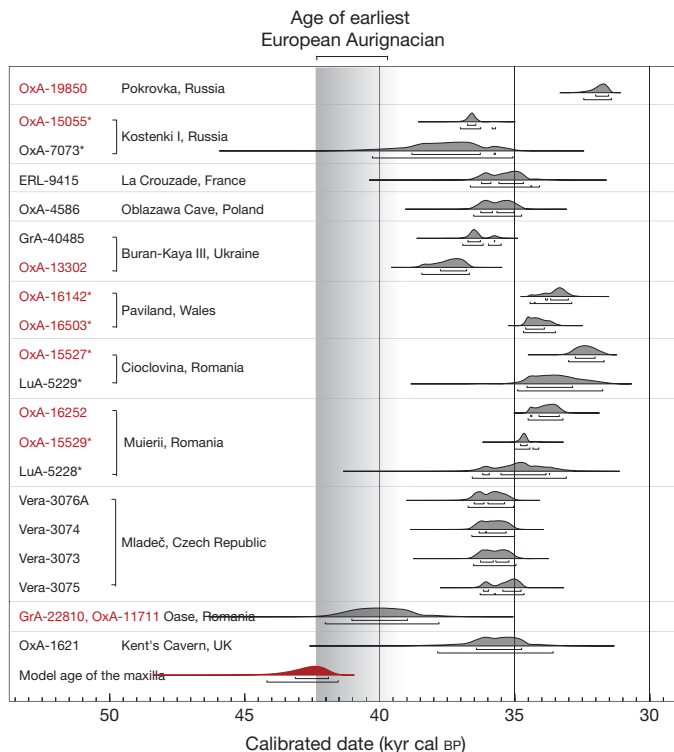
depths between 12–13 ft and 15 ft stratigraphically precede the age of the maxilla and therefore provide a *terminus post quem* for it. Similarly, our modelling of determinations above the maxilla find spot indicate that they stratigraphically follow it in the sequence. These determinations include a date from a cranial fragment of a woolly rhinoceros found just above KC4 (OxA-13965) and three identical dates from a pair of woolly rhinoceros metacarpals found in articulation (Supplementary Methods). These data constrain the probable age of the maxilla and provide a *terminus ante quem* for it. We used the 'date' function in OxCal<sup>11</sup> to calculate a probability distribution function (PDF) for its likely age within this modelled sequence, assuming that it was excavated in its originally deposited location. This PDF corresponds to 43,110–41,890 yr cal BP (68.2% probability) and 44,180–41,530 yr cal BP (95.4% probability)—equivalent to Greenland interstadial 11 and stadial 11 on the NGRIP  $\delta^{18}\text{O}$  Hulu-tuned timescale<sup>13,14</sup> (Supplementary Fig. 2). When we tested the model by varying the priors used, the result for the PDF did not change significantly (Supplementary Methods). We attempted another AMS date from the third premolar (P<sup>3</sup>) of the KC4 specimen to confirm this age, but too little collagen could be extracted for a reliable AMS date to be determined (Supplementary Table 2).

Comparison of the PDF for KC4 with other direct dates of AMH fossils shows it to be earlier than any other (Fig. 3). However, it overlaps with the remains believed to be earliest dated evidence for



**Figure 2 | Bayesian age model for Kent's Cavern.** The model incorporates the radiocarbon likelihoods (calibrated probability distributions; brackets under the distributions represent the 68.2% and 95.4% ranges, respectively) within a relative age sequence based on the depths recorded during the excavation of phases and stalagmite deposits throughout the site. This allows a probability distribution function (PDF) for the estimated age of KC4 (Age of the maxilla PDF; see inset for magnified view) to be generated. Figures in brackets next to OxA numbers are conventional radiocarbon ages. Dates are calibrated using the INTCAL09 curve<sup>12</sup> and compared against the NGRIP  $\delta^{18}\text{O}$  palaeotemperature record<sup>13</sup> tuned to the Hulu Cave timescale<sup>14</sup> ( $\delta^{18}\text{O} = (\delta^{18}\text{O}/$

$\delta^{16}\text{O})_{\text{sample}}/(\delta^{18}\text{O}/\delta^{16}\text{O})_{\text{standard}} - 1$ ). Numbers on the  $\delta^{18}\text{O}$  record represent Greenland interstadials. The three calibrated ages in red are outliers in the model. One of these is the original AMS date of KC4 (OxA-1621), which is shown alongside the new modelled age for the maxilla. It was not included in the Bayesian analysis. The stratigraphic section on the left was adapted from the AR4271 Section Drawing of Vestibule Trench in Ogilvie's excavation archive (Torquay Museum, UK). Location of dated samples by depth is shown: red star, maxilla depth; GS, granular stalagmite, from which depth measurements were recorded; LS, lower stalagmite. Figure generated using OXCAL 4.1<sup>11</sup>. See Supplementary Methods for details of the modelling and testing.



**Figure 3 | Comparison of direct radiocarbon determinations of AMH bones from European Palaeolithic sites<sup>3,9,19–25</sup> with the KC4 model age.** Calibrated using the INTCAL09 curve<sup>12</sup>. Brackets under the distributions represent the 68.2 and 95.4% probability ranges, respectively. The PDF derived from the Bayesian modelling of KC4 (Model age of the maxilla, in red) is earlier than the original direct date from Kent's Cavern (OxA-1621) and all others, and overlaps the start of the age range of the earliest European Aurignacian<sup>15</sup>, which is widely accepted as being linked with the earliest AMH. Ultrafiltered collagen radiocarbon dates are indicated with red text; non-ultrafiltered dates are in black. Asterisks denote duplicate dates on the same human bone. The Oase date is a mean of two determinations, one ultrafiltered and one not<sup>3</sup>.

the Aurignacian of Europe<sup>15</sup> and for the latest Neanderthal and Mousterian sites. Thus, it is crucial to determine whether Keith's attribution<sup>1</sup> of the maxilla as modern human is reliable.

Keith's original illustrations<sup>1</sup> show KC4 essentially as it appears today (Fig. 1A), except for increased cracking due to the specimen drying out and the removal of some alveolar bone for the previous AMS determination. Keith reported that three teeth—the canine, second premolar ( $P^4$ ) and first molar ( $M^1$ )—were in place, whereas the sockets for  $P^3$  and the second molar ( $M^2$ ) were empty. This implies that the specimen had already been reconstructed when Keith received it for study, and may explain why he never questioned the identity of the premolar. He provided basic data on the crown and root dimensions of the three teeth, and the total length of the tooth row as preserved, finding the measurements entirely comparable to those of recent modern humans rather than to those of Neanderthals. However, our study of the specimen shows that  $P^4$  is in fact a right  $P^3$ . To test this re-identification, a virtual three-dimensional model based on a computed microtomography ( $\mu$ CT) scan of the specimen was generated, and alternative positions for the premolar were tested (Supplementary Methods). This revealed a displaced bone fragment in the  $P^3$  socket, and when this was repositioned virtually, the imaged premolar could be rotated and positioned well into this socket (Fig. 1C). Following virtual testing, a new physical reconstruction was then created.

The wear on the teeth of KC4 is such that all occlusal morphology has been removed. Other aspects of dental morphology and measurement were therefore used to determine whether the specimen is Neanderthal or modern human. To facilitate root and pulp chamber

comparisons, computed tomography (CT) was used on modern human fossils from Gough's Cave, UK, and Neanderthal fossils from different sites (Supplementary Methods). The characteristics of each tooth taken individually suggest an identification of Upper Palaeolithic *Homo sapiens*. The premolar is characterized by a low angle of inclination to the vertical of the occlusal part of the buccal surface, small dimensions and a characteristic shape of the pulp chamber and root canals (Fig. 1B), small cervical dimensions and the small size of these relative to the canine and molar, and a narrowness of the root. The canine has a small or absent tuberculum dentale, mild vertical curvature of the buccal side of the root, slight mesial and distal grooves on the root, an absence of buccal and lingual dentine spurs in the root canal, pronounced lingual narrowing of the canal towards the apex and pronounced cervical proportions relative to those of the crown (Fig. 1D). Finally, the molar has a rectangular shape when viewed horizontally, occlusally and at the cervix, and in addition has a less skewed shape in the pulp chamber polygon, an absence of taurodontism, and relative cervical–crown proportions indicative of a modern human (Supplementary Methods).

However, some characteristics of the teeth are ambiguous or more Neanderthal-like. In the canine, there are irregularities in the root canal and the implied presence of tubercle extensions, although the latter are also found in Upper Palaeolithic teeth. Other traits seemingly indicative of non-recent humans are the straight, lingually pointed lingual roots of the first and second molars and the fact that this feature is more pronounced in the second molar than in the first.

Nevertheless, in all but three of 16 dental characteristics KC4 shows AMH rather than Neanderthal affinities, whereas seven remain ambiguous (Table 1 and Supplementary Methods). This suggests strongly that this is an early modern human (*H. sapiens*) fossil. The three possible Neanderthal traits may reflect inadequate sampling of modern human variation, the presence of shared primitive features in KC4 and Neanderthals, or indications of gene flow between Neanderthals and early modern humans<sup>16,17</sup>.

**Table 1 | Dental features of the KC4 maxilla suggesting modern human or Neanderthal features**

Type	No.	Trait
Traits suggesting modern human affinities	1	Canine tuberculum dentale small or absent
	2	Mild vertical convexity of buccal side of canine root
	3	Vertical grooves on canine root shallow and narrow
	4	No buccal or lingual dentine spurs apical to cervix in canine or $P^3$ pulp chambers
	5	Pronounced lingual narrowing of canine root canal towards apex
	6	$P^3$ low angle of inclination of occlusal part of buccal surface relative to vertical axis
	7	Small dimensions and characteristic shape of $P^3$ pulp chamber and root canals
	8	Rectangular shape of $M^1$ occlusally and at cervix
	9	Less skewed shape of $M^1$ pulp chamber polygon
	10	No taurodontism of $M^1$
	11	Low robusticity of $P^3$ root
	12	Small actual and relative cervical dimensions of $P^3$
	13	Relative cervical and crown dimensions of canine and $M^1$
Traits suggesting Neanderthal affinities	14	Possible tubercle extensions on canine
	15	Irregularities in sides of root canals, especially in the canine
Ambiguous traits	16	Greater lingual inclination of $M^2$ lingual root than is found in $M^1$
	17	Crown dimensions and root lengths lie in both Neanderthal and modern human ranges
	18	Canine angle of inclination of occlusal part of buccal surface relative to vertical axis
	19	Mesiobuccal bulge in occlusal outline of canine but not of $P^3$
	20	Canine root surface smooth
	21	Single-rooted $P^3$
	22	Dimensions of canine pulp chamber
	23	Height of $P^3$ pulp chamber roof relative to cervix

Details of the analysis of each trait can be found in Supplementary Methods.

To test further for possible Neanderthal affinities, we attempted the genetic analysis of DNA from the dentine of P<sup>3</sup>, using previously described techniques for isolating and amplifying DNA from preserved remains<sup>18</sup>. PCR amplicons from four positive amplifications were cloned into competent bacterial cells. Analysis of the 72 colonies sequenced from KC4 revealed at least 18 distinct mitochondrial DNA sequences, all of which fell within the range of genetic diversity of AMH. However, given the difficulty of distinguishing between endogenous sequences of modern humans and contamination, we consider these results to be inconclusive regarding the specific identity of KC4. It is not possible to determine whether the specimen is genetically similar to modern humans, or whether the DNA in the sample is so degraded that no endogenous (Neanderthal or otherwise) DNA was recoverable using the methods applied (Supplementary Methods).

The new estimated age for KC4 fills a hitherto unexplained gap between the evidence for the earliest Aurignacian remains and that for the earliest modern human skeletal remains, and adds further strength to the association between the two because of the degree of overlap between the age and dates for the earliest European Aurignacian in western Europe<sup>15</sup>. Although the Upper Palaeolithic artefacts from the deeper parts of the excavations at Kent's Cavern cannot confidently be assigned a precise cultural attribution, we consider an early Aurignacian association for KC4 to be the most likely. Even if this cannot be established, by confirmation of its attribution as an AMH fossil, KC4 demonstrates the very rapid dispersal of early modern humans across Europe, well before 40 kyr ago.

## METHODS SUMMARY

Bone collagen for radiocarbon dating was extracted at the Oxford Radiocarbon Accelerator Unit, University of Oxford, UK, using sequential decalcification, base wash, gelatinization, filtration and ultrafiltration steps to remove low-molecular-mass contaminants. Radiocarbon determinations were obtained using an HVEE accelerator mass spectrometer and calculated conventionally as ages before AD 1950. Calibration was against the INTCAL09 data set, using the OXCAL 4.1 computer program. Bayesian modelling was applied using OXCAL and a relative depth sequence was obtained from the original excavation plans. Outlier detection analysis was used to give less weight to radiocarbon likelihoods in disagreement with the prior information applied. Model sensitivity was tested by varying the priors for the upper parts of the archaeological modelled sequence. New measurements of the morphology of the KC4 teeth were taken from CT scans.  $\mu$ CT scanning of the teeth and maxilla was done using an X-Tek HMX 160 system (X-Tek Systems Ltd). A virtual three-dimensional model of the specimen was created by a combination of thresholding and manual segmentation, and was used for quantitative and qualitative analysis. Comparative dimensions and morphological data were obtained for European Neanderthals, European Upper Palaeolithic *H. sapiens* and recent Europeans from the literature, casts in the collection of the Natural History Museum, London, UK, and computed tomography scans available on the NESPOS website (<https://www.nespos.org/display/openspace/home>). The crown measurements of KC4 made by Keith<sup>1</sup> (corrected) were used for comparison, because these were made on the specimen when it had the fewest cracks. Root robusticity, cervical measurements and internal measurements were the principal source for comparisons, owing to the effects of crown wear on the KC4 teeth. All comparative data are listed in Supplementary Methods.

Received 30 May; accepted 19 August 2011.

Published online 2 November 2011.

1. Keith, A. Report on a fragment of a human jaw. *Trans. Proc. Torquay Nat. Hist. Soc.* **5**, 1–2 (1927).
2. Hedges, R. E. M., Housley, R. A., Law, I. A. & Bronk, C. R. Radiocarbon dates from the Oxford AMS system: *Archaeometry* datelist 9. *Archaeometry* **31**, 207–234 (1989).
3. Trinkaus, E. *et al.* An early modern human from the Peștera cu Oase, Romania. *Proc. Natl Acad. Sci. USA* **100**, 11231–11236 (2003).
4. Lundberg, J. & McFarlane, D. A. Pleistocene depositional history in a periglacial terrane: a 500 ky record from Kents Cavern, Devon, United Kingdom. *Geosphere* **3**, 199–219 (2007).
5. Dowie, H. G. & Ogilvie, A. H. in *Report of the Ninety-Fifth Meeting of the British Association for the Advancement of Science* 303–306 (Murray, 1927).

6. Jacobi, R. M. A collection of Early Upper Palaeolithic artefacts from Beedings, near Pulborough, West Sussex, and the context of similar finds from the British Isles. *Proc. Prehist. Soc.* **73**, 229–325 (2007).
7. Flas, D. La transition du Paléolithique moyen au supérieur dans la plaine septentrionale de l'Europe. *Anthropol. Préhist.* **119**, 5–254 (2008).
8. Jacobi, R. M., Higham, T. F. G. & Bronk Ramsey, C. AMS radiocarbon dating of Middle and Upper Palaeolithic bone in the British Isles: improved reliability using ultrafiltration. *J. Quaternary Sci.* **21**, 557–573 (2006).
9. Higham, T. F. G., Jacobi, R. M. & Bronk Ramsey, C. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**, 179–195 (2006).
10. Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen extraction by modified Longin method. *Radiocarbon* **30**, 171–177 (1988).
11. Bronk Ramsey, C. Development of the radiocarbon calibration program OxCal. *Radiocarbon* **43**, 355–363 (2001).
12. Reimer, P. J. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009).
13. Andersen, K. K. *et al.* The Greenland ice core chronology 2005, 15–42 ka. Part 1: constructing the time scale. *Quat. Sci. Rev.* **25**, 3246–3257 (2006).
14. Weninger, B. & Jöris, O. A <sup>14</sup>C age calibration curve for the last 60 ka: the Greenland-Hulu U/Th timescale and its impact on understanding the Middle to Upper Paleolithic transition in western Eurasia. *J. Hum. Evol.* **55**, 772–781 (2008).
15. Zilhão, J. & d'Errico, F. The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neanderthal extinction. *J. World Prehist.* **13**, 1–68 (1999).
16. Trinkaus, E. European early modern humans and the fate of the Neandertals. *Proc. Natl Acad. Sci. USA* **104**, 7367–7372 (2007).
17. Green, R. E. *et al.* A draft sequence of the Neanderthal genome. *Science* **328**, 710–722 (2010).
18. Svensson, E. M. *et al.* Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Anim. Genet.* **38**, 378–383 (2007).
19. Akimova, E. *et al.* A new direct radiocarbon AMS date for an Upper Palaeolithic human bone from Siberia. *Archaeometry* **52**, 1122–1130 (2010).
20. Henry-Gambier, D. & Sacchi, D. La Crouzade V–VI (Aude, France): un des plus anciens fossiles d'anatomie moderne en Europe occidentale. *Bull. Mem. Soc. Anthropol. Paris* **20**, 79–104 (2008).
21. Prat, S. *et al.* The oldest anatomically modern humans from far southeast Europe: direct dating, culture and behavior. *PLoS ONE* **6**, e20834 (2011).
22. Higham, T. F. G., Bronk Ramsey, C., Brock, F., Baker, D. & Ditchfield, P. Radiocarbon dates from the Oxford AMS system: *Archaeometry* datelist 32. *Archaeometry* **49** (suppl. 1), S1–S60 (2007).
23. Soficaru, A., Petrea, C., Dobos, A. & Trinkaus, E. The human cranium from the Peștera Cioclovina Uscată, Romania. *Curr. Anthropol.* **48**, 611–619 (2007).
24. Soficaru, A., Dobos, A. & Trinkaus, E. Early modern humans from the Peștera Muierii, Baia de Fier, Romania. *Proc. Natl Acad. Sci. USA* **103**, 17196–17201 (2006).
25. Wild, E. *et al.* Direct dating of early Upper Palaeolithic human remains from Mladeč. *Nature* **435**, 332–335 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Kruszynski for allowing us access to fossils and casts in his care and for his help. S. Taft performed CT scans of the specimens, N. Curtis created a preliminary CT-based model and R. Abel provided CT scans of comparative specimens. We thank M. Bradtmöller of NESPOS for his assistance. H. Livsidge provided specimens of modern teeth. S. Bello helped with the Alicona microscope. A. Coppa made available dental morphological data for upper canines, and R. L. Tompkins donated X-ray photographs of fossil dentitions. M. Skinner, P. Gunz, M. Richards, A. Olejniczak and J.-J. Hublin advised on investigative approaches to the study of the specimen. H. Taylor photographed KC4. We are grateful to the following people for making available CT scans of fossils from Neanderthal sites for viewing on the NESPOS database: R. Macchiarelli and J. F. Tournepie (La Chaise de Vouthon), R. Macchiarelli and D. Berthet (La Quina), B. Illerhaus (Le Moustier) and P. Semal (Spy). Funding was provided by the Leverhulme Trust and the NERC. We thank all staff at the Oxford Radiocarbon Accelerator Unit. Funding was provided for T.H. through the NERC (grant NE/D014077/1). C.S. and T.H. are Members, and T.C. is an Associate Member, of the Ancient Human Occupation of Britain project, funded by the Leverhulme Trust.

**Author Contributions** C.S., R.J., E.T. and T.H. initiated the research. C.S., R.J., T.H. and T.C. designed the research. R.J. determined the spatial and depth locations of the AMS-dated bones, identified the material and analysed the site's lithic remains. T.H. did the AMS dating and Bayesian analysis. T.C. analysed the dental traits. C.F. and S.H. provided comparative dental data. B.S. did the DNA analysis. C.C. reconstructed and conserved the maxilla. B.C. curated the KC4 maxilla. P.O'H. and M.F. did the  $\mu$ CT scanning. F.G. performed the virtual three-dimensional reconstruction of KC4 and prepared the Gough's Cave CT scans. F.G. and T.C. made the KC4 dental measurements. T.H., C.S. and T.C. wrote the initial draft of the paper. All authors contributed to and helped to edited the final manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.H. ([thomas.higham@rlaha.ox.ac.uk](mailto:thomas.higham@rlaha.ox.ac.uk)) or C.S. ([c.stringer@nhm.ac.uk](mailto:c.stringer@nhm.ac.uk)).



# Early dispersal of modern humans in Europe and implications for Neanderthal behaviour

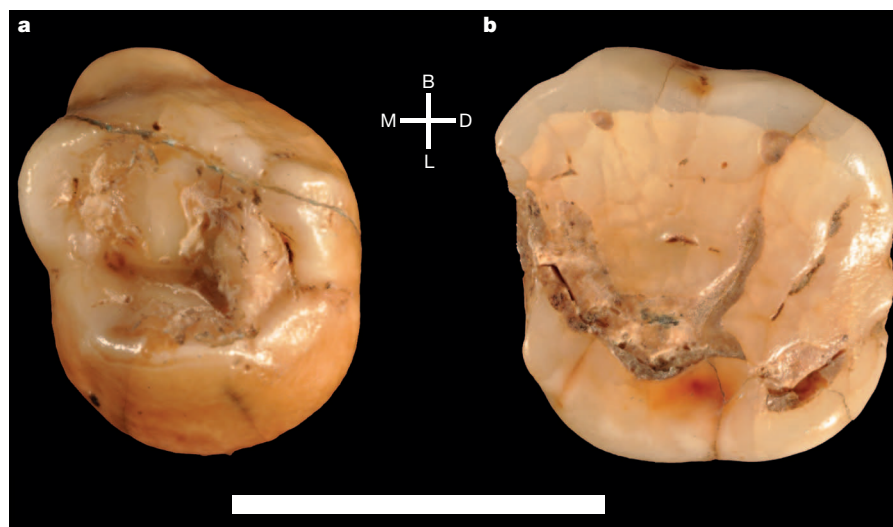
Stefano Benazzi<sup>1</sup>, Katerina Douka<sup>2</sup>, Cinzia Fornai<sup>1</sup>, Catherine C. Bauer<sup>3</sup>, Ottmar Kullmer<sup>4</sup>, Jiří Svoboda<sup>5,6</sup>, Ildikó Pap<sup>7</sup>, Francesco Mallegni<sup>8</sup>, Priscilla Bayle<sup>9</sup>, Michael Coquerelle<sup>10</sup>, Silvana Condemi<sup>11</sup>, Annamaria Ronchitelli<sup>12</sup>, Katerina Harvati<sup>3,13</sup> & Gerhard W. Weber<sup>1</sup>

The appearance of anatomically modern humans in Europe and the nature of the transition from the Middle to Upper Palaeolithic are matters of intense debate. Most researchers accept that before the arrival of anatomically modern humans, Neanderthals had adopted several 'transitional' technocomplexes. Two of these, the Uluzzian of southern Europe and the Châtelperronian of western Europe, are key to current interpretations regarding the timing of arrival of anatomically modern humans in the region and their potential interaction with Neanderthal populations. They are also central to current debates regarding the cognitive abilities of Neanderthals and the reasons behind their extinction<sup>1–6</sup>. However, the actual fossil evidence associated with these assemblages is scant and fragmentary<sup>7–10</sup>, and recent work has questioned the attribution of the Châtelperronian to Neanderthals on the basis of taphonomic mixing and lithic analysis<sup>11,12</sup>. Here we reanalyse the deciduous molars from the Grotta del Cavallo (southern Italy), associated with the Uluzzian and originally classified as Neanderthal<sup>13,14</sup>. Using two independent morphometric methods based on microtomographic data, we show that the Cavallo specimens can be attributed to anatomically modern humans. The secure context of the teeth provides crucial evidence that the makers of the Uluzzian technocomplex

were therefore not Neanderthals. In addition, new chronometric data for the Uluzzian layers of Grotta del Cavallo obtained from associated shell beads and included within a Bayesian age model show that the teeth must date to ~45,000–43,000 calendar years before present. The Cavallo human remains are therefore the oldest known European anatomically modern humans, confirming a rapid dispersal of modern humans across the continent before the Aurignacian and the disappearance of Neanderthals.

Two deciduous molars (Cavallo-B and Cavallo-C) were excavated in 1964 from the site of Grotta del Cavallo (Apulia, southern Italy; Supplementary Information). Cavallo is important as the type site of the Uluzzian technocomplex<sup>15</sup>, one of the three main transitional industries alongside the Châtelperronian and Szeletian, in Franco-Cantabria and Central Europe, respectively. These are strongly suspected of being produced by Neanderthals, although the actual fossil evidence in association is scant<sup>16</sup>.

Cavallo-B is a deciduous left upper first molar (dM<sup>1</sup>), found in layer EIII (archaic Uluzzian). Cavallo-C is a deciduous left upper second molar (dM<sup>2</sup>) found 15–20 cm above Cavallo-B, in layer EII-I (evolved Uluzzian)<sup>13</sup> (Supplementary Fig. 1 and Supplementary Table 1). The specimens (Fig. 1) were described in 1967 by Palma di Cesnola and



**Figure 1** | Occlusal view of the deciduous molars from the Uluzzian layers of Grotta del Cavallo (Apulia, southern Italy). **a**, Cavallo-B (deciduous left upper first molar; dM<sup>1</sup>). **b**, Cavallo-C (deciduous left upper second molar; dM<sup>2</sup>). B, buccal; D, distal; L, lingual; M, mesial. Scale bar, 1 cm.

<sup>1</sup>Department of Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. <sup>2</sup>Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK. <sup>3</sup>Paleoanthropology, Department of Early Prehistory and Quaternary Ecology, Eberhard Karls Universität Tübingen, Rümelnstrasse 23, Tübingen 72070, Germany. <sup>4</sup>Department of Paleoanthropology and Messel Research, Senckenberg Research Institute Frankfurt, Senckenberganlage 25, D-60325 Frankfurt, Germany. <sup>5</sup>Institute of Archaeology, Academy of Sciences of the Czech Republic, Královopolská 147, 612 00 Brno, Czech Republic. <sup>6</sup>Department of Anthropology, Faculty of Science, Masaryk University, Viniřská 5, 603 00 Brno, Czech Republic. <sup>7</sup>Department of Anthropology, Hungarian Natural History Museum, Ludovika tér 2-6, 1083 Budapest, Hungary. <sup>8</sup>Department of Biology, University of Pisa, Via S. Maria 53, 56126 Pisa, Italy. <sup>9</sup>UMR 5199 PACEA, Université Bordeaux 1, avenue des Facultés, 33405 Talence, France. <sup>10</sup>Paleoanthropology group, Department of Paleobiology Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain. <sup>11</sup>UMR 6578 CNRS/Aix Marseille/EFS, Laboratoire d'Anthropologie Bioculturelle, Faculté de Médecine/Secteur Nord, CS80011 Bd Pierre Dramard 13344, Marseille Cedex 15, France. <sup>12</sup>Department of Environmental Sciences "G. Sarfatti", U.R. Prehistoric Ecology, University of Siena, via T. Pendola 62, 53100 Siena, Italy. <sup>13</sup>Senckenberg Center for Human Evolution and Paleoecology, Eberhard Karls Universität Tübingen, Rümelnstrasse 23, Tübingen 72070, Germany.

Messeri<sup>13</sup> who classified Cavallo-B as modern human and Cavallo-C as Neanderthal. On this basis, the authors suggested a persistence of Neanderthal populations in southern Italy after the appearance of modern humans<sup>17</sup>.

Although information about these specimens is scarce and contradictory, most scholars accept that the deciduous molars from Grotta del Cavallo are attributable to Neanderthals, and therefore that Neanderthals produced the Uluzzian. This attribution was proposed by ref. 14 for Cavallo-B on the basis of the specimen's crown diameters (the dimensions of Cavallo-C were found to be compatible with both Neanderthals and anatomically modern humans). However, the Cavallo-B buccolingual and mesiodistal diameters used by ref. 14 appear to have been accidentally substituted for each other when compared to the crown diameters reported by ref. 13. The correct crown diameter values do not support Neanderthal affinities for Cavallo-B, neither does the crown morphology of the two specimens. Cavallo-B shows three dental cusps, typical of  $dm^1$  (ref. 10) from anatomically modern humans, with the lingual cusps mesially oriented and separated from the buccal cusps by a well-defined sagittal sulcus. Conversely, the  $dm^1$  of Neanderthals is more frequently four-cusped, with cusp tips compressed internally<sup>10</sup>. Cavallo-C has a sub-square crown outline, similar to  $dm^2$  of anatomically modern humans and different from the typical rhomboid outline with distolingual hypocone expansion of Neanderthal  $dm^2$ s (ref. 10).

To establish firmly the taxonomic affinities of the Cavallo human remains, we re-analysed Cavallo-B and Cavallo-C with two independent morphometric methods, using a comparative sample of Neanderthal, Upper Palaeolithic modern human (UPMH) and recent modern human (RMH)  $dm^1$  and  $dm^2$  specimens (Supplementary Tables 2–4).

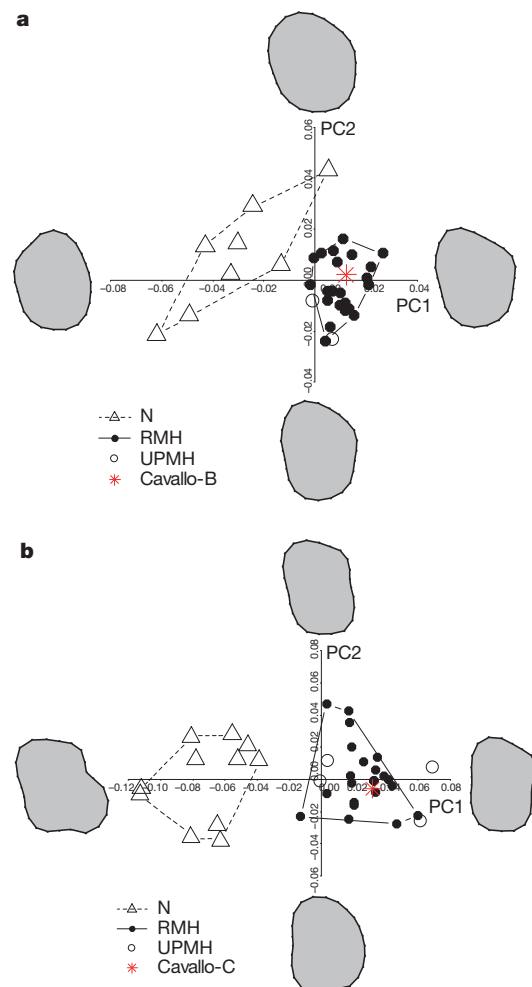
Our first approach is a geometric morphometric analysis of outlines obtained from the dental crown<sup>18</sup> (see Methods), for which the group shape variation was evaluated through a shape-space principal component analysis (PCA).

For the  $dm^1$  crown outlines (Fig. 2a), the first two principal components (PCs) account for about 63% of the total variance. Neanderthals and modern humans separate along PC1 (42.7%), which characterizes size-independent shape variation ( $r = -0.33$ ;  $P = 0.06$ ). Neanderthal  $dm^1$  specimens show an ovoid outline, whereas RMH and UPMH specimens are more irregularly shaped for the presence of well-expressed tuberculum molare (molar tubercle of Zuckerkandl) and metacone (buccodistal) cusp, and for a general distolingual constriction due to the reduction of the hypocone. Cavallo-B plots well within the range of variability of the anatomically modern human sample. The cross-validation quadratic discriminant analysis (QDA) of the PC1 scores classified Cavallo-B as modern human with a posterior probability ( $P_{\text{post}}$ )  $> 0.90$  (Supplementary Table 4).

In the analysis of the  $dm^2$  cervical outlines (Fig. 2b), the first two PCs account for about 84% of the total variance. Neanderthals and anatomically modern humans are even more clearly separated along PC1 (71.4%), which expresses size-dependent shape variation (static allometry,  $r = 0.74$ ;  $P < 0.001$ ). Neanderthal  $dm^2$  specimens are characterized by a rhomboid cervical outline due to their large hypocone, whereas UPMH and RMH specimens have sub-square outlines. Cavallo-C plots unambiguously within the modern human range. The cross-validation QDA of the first two PC scores attributes Cavallo-C to modern human with a  $P_{\text{post}} > 0.90$  (Supplementary Table 4).

The second morphometric method considers the internal structure of the teeth and consists of the two-dimensional enamel thickness and dental tissue proportions analysis (Fig. 3) (see Methods and Supplementary Tables 3 and 4). The average and relative enamel thickness (AET and RET, respectively) have been described as effective taxonomic discriminators between Neanderthals and modern humans because Neanderthal molars are characterized by significantly thinner enamel relative to dentine volume<sup>19</sup>.

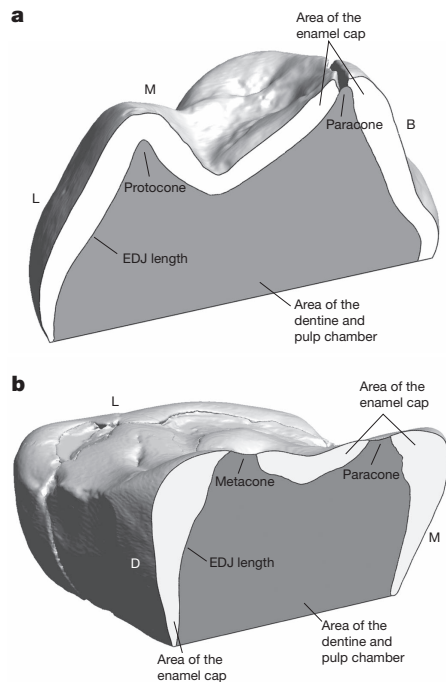
The  $dm^1$  modern human samples (UPMH and RMH) shown in Table 1 have been divided into sub-groups on the basis of their degree



**Figure 2 | Shape-space PCA plots of  $dm^1$  crown outlines and  $dm^2$  cervical outlines. a,  $dm^1$  crown outline. b,  $dm^2$  cervical outline. The deformed mean crown outline in the direction of the PC is drawn at the extremity of each axis. N, Neanderthal; RMH, recent modern human; UPMH, Upper Palaeolithic modern human.**

of wear (unworn/wear stage 1 distinguished by wear stage 3; based on Smith<sup>20</sup>) to ease the comparison with the Neanderthal  $dm^1$  samples, which is entirely affected by wear stage 3. The Neanderthal  $dm^1$  RET indexes are significantly lower than those of RMH at similar wear stages ( $P < 0.001$ ; permutation test,  $n = 1,000$ ) on group mean and variance differences. The AET and RET indexes of Cavallo-B lie beyond the highest values computed so far for the unworn UPMH and RMH (Table 1). Considering that the RET index average difference from unworn to wear stage 3 for both RMH and UPMH is about 0.80, it is reasonable to assume that if Cavallo-B would have worn down to a wear stage 3, it would provide a RET index of approximately 11. This value is still completely outside the Neanderthal range of variation and near the highest values computed for the unworn RMH. The result further supports strongly the affiliation of Cavallo-B as anatomically modern human rather than Neanderthal.

With regard to the  $dm^2$  specimens (Table 1), the Neanderthal RET indexes are significantly lower than those of RMH ( $P < 0.001$ ). The UPMH specimens present the highest AET and RET indexes of the whole sample, RMH included. Cavallo-C is the most worn specimen within our  $dm^2$  sample (wear stage 5), therefore the AET and RET indexes result in a rather lower value than could be expected for the unworn stage of the same tooth. Nonetheless, both indexes still rank among the highest values obtained (Table 1). The cross-validation



**Figure 3 | Cross-sections of Cavallo-B and Cavallo-C for two-dimensional enamel thickness analysis.** **a**, Buccolingual cross-section of Cavallo-B through the dentine horns of the protocone and paracone. **b**, Mesiodistal cross-section of Cavallo-C through the dentine horns of the paracone and metacone. B, buccal; D, distal; EDJ, enamel–dentin junction; L, lingual; M, mesial.

QDA of the  $dM^2$  RET index classifies Cavallo-C as modern human with a  $P_{\text{post}} > 0.90$  (Supplementary Table 4).

New radiocarbon dating of the Uluzzian layers was undertaken to produce a more robust chronology. Previous dates from the site disclosed inconsistency due to incomplete decontamination and unsuitability of the dated samples (Supplementary Information). In the absence of collagen from bone at the site and the lack of charcoal samples collected at the time of the excavation, marine shell samples were the only alternative and were therefore selected for dating. In the Uluzzian layers of Cavallo, several shells were transformed into beads, by snapping or piercing to produce personal ornaments. These beads are generally held to be an indicator of symbolic and complex behaviour. Eight shells of *Dentalium* sp., *Nuculana* sp. and *Cyclope neritea* were dated by accelerator mass spectrometry (AMS) radiocarbon dating following a novel methodological approach (Supplementary Information and Supplementary Fig. 3). The new dates were incorporated into a Bayesian model using the OxCal program (Supplementary Information) and calibrated against the INTCAL09 calibration curve<sup>21</sup> (Supplementary Figs 4 and 5 and Supplementary Tables 5 and 6). Layer EIII was calculated by the model to date between 45,010–43,380

(68.2% probability) and 47,530–43,000 (95.4% probability) calendar years before present (where present = 1950) (cal. yr BP). The distribution falls within Greenland Interstadial (GIS) 12, a long warm phase following Heinrich Event 5, and most likely towards its latter part. A decrease in temperature has been inferred based on the faunal assemblage from EIII (ref. 22). A likely initial arrival of anatomically modern humans during this post-HE5 interstadial has been suggested previously<sup>23</sup>. Layer EII-I, associated with a shell date of 40,000  $^{14}\text{C}$  yr BP, was modelled to date between 44,000–43,000 cal. yr BP (68.2% probability), a similar age to layer EIII. Comparable chronometric results were obtained from Grotta di Fumane, another Uluzzian site in the Italian pre-Alps and the only other with reliable chronometric information, where the technocomplex is dated at 44,800–43,900 cal. yr BP (95.4% probability) (or  $\sim 42,000$ –40,000  $^{14}\text{C}$  yr BP)<sup>24</sup>.

The new chronometric results show that the two deciduous molars from Grotta del Cavallo are the earliest European anatomically modern human fossils currently known. Because the Uluzzian technocomplex stratigraphically underlies the earliest Aurignacian in all instances where the two co-occur (for example, Grotta di Castelcivita, Grotta della Cala, Grotta La Fabbria and Grotta di Fumane)<sup>5</sup>, the arrival of the earliest modern humans at these sites must pre-date the Aurignacian. Furthermore, considering that Neanderthals are likely to have survived in most of continental Europe until at least  $\sim 40,000$  cal. yr BP (ref. 25), our results offer fossil evidence for a longer period of co-existence in Europe between Neanderthals and modern humans.

The re-attribution of the teeth of Grotta del Cavallo to anatomically modern human has implications for the interpretation of the Uluzzian technocomplex<sup>14,16</sup>. The presence of personal ornaments in the form of marine shell beads, worked bone and colorants—including ochre and limonites—in the Uluzzian layers of Grotta del Cavallo<sup>5,6</sup> has been used as direct evidence for Neanderthals reaching behavioural modernity independent of, and before, anatomically modern humans reaching Europe<sup>1,26</sup>. These attributes are all more typical of Upper Palaeolithic industries. This multiple species model for the origin of fully modern behaviour has been considered by some to be an impossible coincidence<sup>27</sup> and a fervent debate has ensued among prehistorians on the behavioural and cognitive capabilities of the makers of the transitional industries found across Europe and the Levant. Our results show that the Uluzzian is not a Neanderthal industry.

Stratigraphically, the Uluzzian is always separated from the final Mousterian by sterile layers, volcanic ash (as in Cavallo), erosional discontinuities or depositional hiatuses, which might suggest that a period of time has elapsed between the two phases. In southern Italy, economic and cultural behaviour of the Uluzzian suggests a greater affinity with the succeeding Aurignacian (with marginally backed tools) than with the final Mousterian<sup>5,6,22</sup>. These findings provide additional support for a modern human authorship of the Uluzzian. Although we cannot extrapolate our conclusions to other transitional industries, our findings indicate that caution should be applied when associating Neanderthals with them, particularly the Châtelperronian and Szeletian (see details of the ongoing debate on this topic<sup>2–4,9,11,12</sup>).

**Table 1 | Two-dimensional enamel thickness of Cavallo-B and Cavallo-C compared with the indicated  $dM^1$  and  $dM^2$  samples.**

Tooth	Taxon	Wear stage*	n	AET (mm)		RET (scale free)	
				Mean	Range	Mean	Range
$dM^1$	Neanderthal	3	6	0.40 (0.03)	0.37–0.45	7.17 (0.54)	6.61–7.93
$dM^1$	UPMH	3	2	0.51 (0.01)	0.50–0.52	9.56 (0.13)	9.47–9.66
$dM^1$	RMH	3	14	0.47 (0.03)	0.43–0.52	9.12 (0.67)	8.50–10.52
$dM^1$	UPMH	Unworn	1	0.56	–	10.36	–
$dM^1$	RMH	Unworn–stage 1	8	0.51 (0.06)	0.41–0.58	9.96 (0.96)	8.66–11.36
$dM^1$	Cavallo-B	Unworn	1	0.69	–	11.80	–
$dM^2$	Neanderthal	Unworn–stage 1	9	0.63 (0.04)	0.58–0.69	10.89 (0.84)	9.60–12.39
$dM^2$	UPMH	Stage 1–stage 2	2	0.97 (0.15)	0.86–1.07	17.93 (1.40)	16.94–18.92
$dM^2$	RMH	Unworn–stage 3	23	0.73 (0.08)	0.56–0.93	13.84 (1.53)	11.43–18.00
$dM^2$	Cavallo-C	Stage 5	1	0.84	–	14.28	–

Standard deviation is indicated in brackets. AET, average enamel thickness index; RET, relative enamel thickness index; RMH, recent modern human; UPMH, Upper Palaeolithic modern human.

\*Based on ref. 20.



The association of the Uluzzian with anatomically modern humans implies much greater complexity and age depth to the movement of modern humans into Europe and may lend support to a southern Mediterranean route in their dispersal, similar to that identified previously<sup>27</sup> for the spread of the Aurignacian. Although it is during the Aurignacian that certain technological and behavioural innovations effloresce, such as blade and bladelet-dominated lithic assemblages, bone and ivory tools, art and personal ornaments, the initial appearance of these traits in southern Europe clearly pre-dates this. This discovery has significant implications for our understanding of the earliest presence of anatomically modern humans in Europe, expands the period of overlap between modern humans and Neanderthals and makes it much less likely that Neanderthals developed their own Upper Palaeolithic suite of behaviours before the arrival of anatomically modern humans.

## METHODS SUMMARY

The comparative dental sample for both the morphometric outline analyses and the two-dimensional enamel thickness and dental tissue proportions analysis is provided in Supplementary Table 2.

Scans of all the specimens were undertaken by means of industrial and synchrotron-based microtomographic scanners at isotropic voxel length between 15 and 55 µm. The microtomographic image stacks of each tooth were aligned with the cervical plane parallel to the *x-y* plane of the Cartesian coordinate system. The three-dimensional digital surface models were created semi-automatically by threshold-based segmentation, contour extraction and surface reconstruction.

For the outline analyses we considered the dM<sup>1</sup> crown outlines because Cavallo-B is unworn; conversely, we used the cervical outlines of the dM<sup>2</sup> samples, as Cavallo-C shows both occlusal and interproximal wear. To identify the crown outline and the cervical outline we followed the procedures described previously<sup>18</sup>, with some adjustment for our specific case.

For the analyses of the two-dimensional enamel thickness and dental tissue proportions, the following measurements were recorded: the area of the enamel cap (mm<sup>2</sup>), the area of the coronal dentine (which includes the coronal pulp; mm<sup>2</sup>), the length of the enamel–dentine junction (EDJ; mm), the average enamel thickness (AET) index (the area of the enamel cap divided by the length of the EDJ; index in mm) and the RET index (the average enamel thickness divided by the square root of the coronal dentine area; scale free index)<sup>19,28</sup>. The data were analysed via software routines written in R<sup>29</sup> (Supplementary Information).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 August; accepted 5 October 2011.

Published online 2 November 2011.

1. d'Errico, F., Zilhão, J., Julien, M., Baffier, D. & Pélegrin, J. Neanderthal acculturation in Western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* **39**, 1–44 (1998).
2. Gravina, B., Mellars, P. & Ramsey, C. B. Radiocarbon dating of interstratified Neanderthal and early modern human occupations at the Châtelperronian type-site. *Nature* **438**, 51–56 (2005).
3. Mellars, P., Gravina, B. & Bronk Ramsey, C. Confirmation of Neanderthal/modern human interstratification at the Châtelperronian type-site. *Proc. Natl Acad. Sci. USA* **104**, 3657–3662 (2006).
4. Zilhão, J. *et al.* Analysis of Aurignacian interstratification at the Châtelperronian-type site and implications for the behavioral modernity of Neandertals. *Proc. Natl Acad. Sci. USA* **103**, 12643–12648 (2006).
5. Ronchitelli, A., Boscato, P. & Gambassini, P. in *La lunga storia di Neandertal. Biologia e comportamento* (eds Facchini, F. & Belcastro, G. M.) (Jaca Book, 2009).
6. d'Errico, F., Borgia, V. & Ronchitelli, A. Uluzzian bone technology and its implications for the origin of behavioural modernity. *Quat. Int.* doi:10.1016/j.quaint.2011.03.039 (in the press).
7. Lévêque, F. & Vandermeersch, B. M. Découverte de restes humains dans un niveau castelperronien à Saint-Césaire (Charente-Maritime). *CR Acad. Sci. Paris* **291**, 187–189 (1980).
8. Hublin, J. J., Spoor, F., Braun, M., Zonneveld, F. & Condemi, S. A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature* **381**, 224–226 (1996).
9. Harvati, K., Panagopoulou, E. & Karkanas, P. First Neanderthal remains from Greece: the evidence from Lakonis. *J. Hum. Evol.* **45**, 465–473 (2003).
10. Bailey, S. E. & Hublin, J. J. Dental remains from the Grotte du Renne at Arcy-sur-Cure (Yonne). *J. Hum. Evol.* **50**, 485–508 (2006).
11. Bar-Yosef, O. & Bordes, J. G. Who were the makers of the Châtelperronian culture? *J. Hum. Evol.* **59**, 586–593 (2010).

12. Higham, T. *et al.* Chronology of the Grotte du Renne (France) and implications for the context of ornaments and human remains within the Châtelperronian. *Proc. Natl Acad. Sci. USA* **107**, 20234–20239 (2010).
13. Palma di Cesnola, A. & Messeri, M. P. Quatre dents humaines paléolithiques trouvées dans des cavernes de l'Italie Méridionale. *L'Anthropologie* **71**, 249–262 (1967).
14. Churchill, S. E. & Smith, F. H. Makers of the early Aurignacian of Europe. *Am. J. Phys. Anthropol.* **113** (Suppl. 31), 61–115 (2000).
15. Palma di Cesnola, A. L'Uluzzien: faciès italien du leptolithique archaïque. *L'Anthropologie* **93**, 783–812 (1989).
16. Riel-Salvatore, J. in *Sourcebook of Paleolithic Transitions* (eds Camps, M. & Chauhan, P.) (Springer, 2009).
17. Messeri, P. & Palma di Cesnola, A. Contemporaneità di paleantropi e fanerantropi sulle coste dell'Italia meridionale. *Zephyrus* **26–27**, 7–30 (1976).
18. Benazzi, S. *et al.* Comparison of dental measurement systems for taxonomic assignment of first molars. *Am. J. Phys. Anthropol.* **144**, 342–354 (2011).
19. Olejniczak, A. J. *et al.* Dental tissue proportions and enamel thickness in Neandertal and modern human molars. *J. Hum. Evol.* **55**, 12–23 (2008).
20. Smith, B. H. Patterns of molar wear in hunter-gatherers and agriculturists. *Am. J. Phys. Anthropol.* **63**, 39–56 (1984).
21. Reimer, P. J. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009).
22. Boscato, P. & Crezzini, J. Middle-Upper Palaeolithic transition in Southern Italy: Uluzzian macromammals from Grotta del Cavallo (Apulia). *Quat. Int.* doi:10.1016/j.quaint.2011.03.028 (in the press).
23. Müller, U. C. *et al.* The role of climate in the spread of modern humans into Europe. *Quat. Sci. Rev.* **30**, 273–279 (2011).
24. Higham, T. European Middle and Upper Palaeolithic radiocarbon dates are often older than they look: problems with previous dates and some remedies. *Antiquity* **85**, 235–249 (2011).
25. Pinhasi, R., Higham, T. F., Golovanova, L. V. & Doronichev, V. B. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc. Natl Acad. Sci. USA* **108**, 8611–8616 (2011).
26. Zilhão, J. in *Continuity and Discontinuity in the Peopling of Europe* (eds Condemi, S. & Weniger, G. C.) (Springer, 2011).
27. Mellars, P. The impossible coincidence: a single species model for the origins of modern human behavior in Europe. *Evol. Anthropol.* **14**, 12–27 (2005).
28. Martin, L. B. Significance of enamel thickness in hominoid evolution. *Nature* **314**, 260–263 (1985).
29. R Development Core Team. R: a language and environment for statistical computing. (<http://www.r-project.org>) (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the Soprintendenza per i Beni Archeologici della Puglia which facilitated the excavation of Grotta del Cavallo over the years. We also thank M. A. Gorgoglione who supported and helped in the collection of samples for <sup>14</sup>C dating and encouraged the collaboration with the University of Siena for the study of the archaeological remains. P. Boscato, H. Klempereva, F. Ranaldo and S. Ricci have all helped in aspects of the research and are especially thanked. We are grateful to G. Gruppioni for providing the Italian modern human sample used in this work. We thank M. Francken, B. Trautmann, I. Trautmann, H. Scherf, M. Dockner and R. Ginner for technical assistance. We thank F. L. Bookstein for suggestions on statistics. Access to the fossil specimens was made possible by the Croatian National History Museum, the French Musée National de Préhistoire, the French Muséum National d'Histoire Naturelle, Paleoanthropology, Eberhard Karls Universität Tübingen and the NESPOS Database 2011 (<https://www.nespos.org/display/openspace/Home>). We acknowledge the Centre de Microtomographie (Université de Poitiers), the Vienna micro-CT Laboratory (University of Vienna), VISCOM AG Hannover, the Paleoanthropology High Resolution Computing Tomography Laboratory (Eberhard Karls Universität Tübingen), the European Synchrotron Radiation Facility beamline ID17, the AST-RX platform (French Muséum National d'Histoire Naturelle) and the Oxford Radiocarbon Accelerator Unit (ORAU). The authors would like to thank T. Higham and R. E. M. Hedges for their input in the radiocarbon dating part of the project, for important comments and proofreading this manuscript. The radiocarbon dating was funded by the Natural Environment Research Council (NERC) NRCF programme. K.D. is part of the Ancient Human Occupation of Britain project, funded by the Leverhulme Trust. This work was supported by the NSF 01-120 Hominid Grant 2007, A.E.R.S. Dental Medicine Organisations GmbH FA547013, the Fondation Fyssen, the DFG INST 37/706-1 FUGG and the NERC Grant (NE/D014077/1).

**Author Contributions** S.B., F.M., K.H. and G.W.W. initiated and organized the project. S.B., C.C.B., P.B., J.S., I.P., K.H. and G.W.W. collected the fossils and modern human sample. S.B. and C.F. carried out the dental measurements. S.B. and M.C. analysed the data. K.D. initiated and performed the radiocarbon dating project. S.B., K.D., C.F., O.K., M.C., S.C., A.R., K.H. and G.W.W. discussed the results. S.B., K.D., C.F., M.C., S.C., A.R., K.H. and G.W.W. wrote and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.B. ([stefano.benazzi@univie.ac.at](mailto:stefano.benazzi@univie.ac.at)).

## METHODS

### Scanning, segmentation and three-dimensional reconstruction of the specimens.

Microtomographic scans of all the specimens (Supplementary Table 2) were undertaken by means of industrial and synchrotron-based microtomographic scanners at isotropic voxel length between 15 and 55  $\mu\text{m}$ . The microtomographic image stacks of each tooth were aligned to the best-fit plane computed at the cervical line (cervical plane) through Amira 5.3 (Mercury Computer Systems), and then rotated up so that the cervical plane was parallel to the  $x$ - $y$  plane of the Cartesian coordinate system. For the segmentation process, the half-maximum height (HMH) protocol was used to reconstruct three-dimensional digital surface models of each microtomography-scanned tooth using Amira 5.3 (Supplementary Information).

**Outline data.** A further orientation of the  $\text{dM}^1$  and  $\text{dM}^2$  digital models in the Cartesian coordinate system was required before the outline analysis. The  $\text{dM}^1$  oriented digital models were rotated around the  $z$  axis to align the projection on the  $x$ - $y$  plane of the intercept between the paracone and protocone cusp tips parallel to the  $y$  axis of the Cartesian coordinate system. The crown outlines were then projected onto the  $x$ - $y$  plane. The  $\text{dM}^2$  digital models were oriented with the lingual side parallel to the  $x$  axis. The best-fit plane of the cervical line identified the cervical outline. In one case (the Neanderthal specimen Engis 2), computer-aided design (CAD) techniques were used to restore the  $\text{dM}^1$  crown outline (Supplementary Fig. 2).

All the outlines were represented by 24 landmarks obtained by equiangularly spaced radial vectors out of the centroid of their area<sup>18</sup>. These landmarks were superimposed through a generalized Procrustes analysis (GPA)<sup>30</sup>. Since the outlines were oriented and centred on the centroid of their area<sup>18</sup>, GPA only entailed a uniform scaling of the landmark configurations to unit centroid size. This step removed size differences, except for static allometry.

**Two-dimensional enamel thickness data.** For the two-dimensional enamel thickness assessment we followed techniques developed previously<sup>28</sup>, adapted to our specific case. A plane perpendicular to the cervical plane of the tooth and passing through two dentine horn tips was used for sectioning the dental crowns. The section passed through the  $\text{dM}^1$ 's paracone and protocone dentine horn tips, and through the  $\text{dM}^2$ 's paracone and metacone dentine horn tips (Fig. 3). The dentine horn tips were identified as the highest points of the dentine in the central mammelon by scrolling apically through the oriented slices. For Cavallo-C (wear stage 5; based on Smith<sup>20</sup>) and the comparative sample with wear stage 3, this approach was further verified by segmenting the whole crown dentine to check the continuity of the marginal ridges beside the dentine horns. For worn teeth, the EDJ length was truncated at the exposed edge of the occlusal dentine basins.

Cavallo-B shows a crack crossing the paracone mesiodistally (Fig. 3a). The area of the crack pertaining to the dentine was included in the dentine area, and the trait

of the EDJ intersected by the crack was calculated in the EDJ length. On the contrary, the missing area of the enamel cap was not reconstructed. Therefore, the computed RET index for Cavallo-B is slightly underestimated.

The segmentation process and the parameter measurements were carried out by S.B. and C.F. The interobserver error was evaluated for the enamel and dentine area of three of the fossil specimens from our sample, and did not exceed 3% in each case.

The following measurements were recorded: the area of the enamel cap ( $\text{mm}^2$ ), the area of the coronal dentine (which includes the coronal pulp;  $\text{mm}^2$ ), the length of the enamel-dentine junction (EDJ; mm), the average enamel thickness (AET) index (the area of the enamel cap divided by the length of the EDJ; index in mm) and the RET index (the average enamel thickness divided by the square root of the coronal dentine area; scale free index)<sup>19,28</sup>.

**Statistical analysis.** A Principal component analysis (PCA) of the matrix of Procrustes coordinates was carried out for the  $\text{dM}^1$  crown outlines and  $\text{dM}^2$  cervical outlines, separately. Because we aimed at assessing the dental outline shapes of the teeth from Grotta del Cavallo with respect to both the anatomically modern human and the Neanderthal outline shape variation, Cavallo-B and Cavallo-C spatial configurations were projected into the space built from the comparative sample only. In the Supplementary Information we show that the first PC of  $\text{dM}^1$  crown outlines PCA and the first two PCs of  $\text{dM}^2$  cervical outlines PCA are the only informative ones.

We did not regress the part of radial size correlated with the diameters, diagonals and area out of the crown and cervical outline data (as previously suggested<sup>18</sup> for the first permanent molars) because size information related to static allometry is an important factor for the separation of Neanderthal and anatomically modern human  $\text{dM}^2$ s.

The differences between the AET and RET indexes of Neanderthal and RMH were tested via a permutation test ( $n = 1,000$ ) on group mean and variance.

Finally, we used leave-one-out cross-validation quadratic discriminant analysis (QDA) for the taxonomic classification of Cavallo-B and Cavallo-C. We built up QDA models leaving out the data from the Cavallo specimens. The computation of the posterior probabilities ( $P_{\text{post}}$ ) was made with an equal prior probability ( $P_{\text{prior}}$ ) of 0.5 for Neanderthal and anatomically modern human groups (UPMH plus RMH). The threshold for taxonomic determination was a  $P_{\text{post}} \geq 0.90$ . The taxonomic analyses are summarized in the Supplementary Information. The data were processed and analysed through software routines written in R<sup>29</sup>.

30. Rohlf, F. J. & Slice, D. E. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* **39**, 40–59 (1990).

# Ascaris suum draft genome

Aaron R. Jex<sup>1\*</sup>, Shiping Liu<sup>2\*</sup>, Bo Li<sup>2\*</sup>, Neil D. Young<sup>1\*</sup>, Ross S. Hall<sup>1</sup>, Yingrui Li<sup>2</sup>, Linfeng Yang<sup>2</sup>, Na Zeng<sup>2</sup>, Xun Xu<sup>2</sup>, Zijun Xiong<sup>2</sup>, Fangyuan Chen<sup>2</sup>, Xuan Wu<sup>2</sup>, Guojie Zhang<sup>2</sup>, Xiaodong Fang<sup>2</sup>, Yi Kang<sup>2</sup>, Garry A. Anderson<sup>1</sup>, Todd W. Harris<sup>3</sup>, Bronwyn E. Campbell<sup>1</sup>, Johnny Vlaminck<sup>4</sup>, Tao Wang<sup>4</sup>, Cinzia Cantacessi<sup>1</sup>, Erich M. Schwarz<sup>5</sup>, Shoba Ranganathan<sup>6</sup>, Peter Geldhof<sup>4</sup>, Peter Nejsun<sup>7</sup>, Paul W. Sternberg<sup>5</sup>, Huanming Yang<sup>2</sup>, Jun Wang<sup>2</sup>, Jian Wang<sup>2</sup> & Robin B. Gasser<sup>1</sup>

Parasitic diseases have a devastating, long-term impact on human health, welfare and food production worldwide. More than two billion people are infected with geohelminths, including the roundworms *Ascaris* (common roundworm), *Necator* and *Ancylostoma* (hookworms), and *Trichuris* (whipworm), mainly in developing or impoverished nations of Asia, Africa and Latin America<sup>1</sup>. In humans, the diseases caused by these parasites result in about 135,000 deaths annually, with a global burden comparable with that of malaria or tuberculosis in disability-adjusted life years<sup>1</sup>. *Ascaris* alone infects around 1.2 billion people and, in children, causes nutritional deficiency, impaired physical and cognitive development and, in severe cases, death<sup>2</sup>. *Ascaris* also causes major production losses in pigs owing to reduced growth, failure to thrive and mortality<sup>2</sup>. The *Ascaris*-swine model makes it possible to study the parasite, its relationship with the host, and ascariasis at the molecular level. To enable such molecular studies, we report the 273 megabase draft genome of *Ascaris suum* and compare it with other nematode genomes. This genome has low repeat content (4.4%) and encodes about 18,500 protein-coding genes. Notably, the *A. suum* secretome (about 750 molecules) is rich in peptidases linked to the penetration and degradation of host tissues, and an assemblage of molecules likely to modulate or evade host immune responses. This genome provides a comprehensive resource to the scientific community and underpins the development of new and urgently needed interventions (drugs, vaccines and diagnostic tests) against ascariasis and other nematodiasis.

We sequenced the *A. suum* genome at ~80-fold coverage (Supplementary Fig. 1), producing a final draft assembly of 272,782,664 base pairs (bp) (N50 = 407 kilobases, kb; N90 = 80 kb; 1,618 contigs of

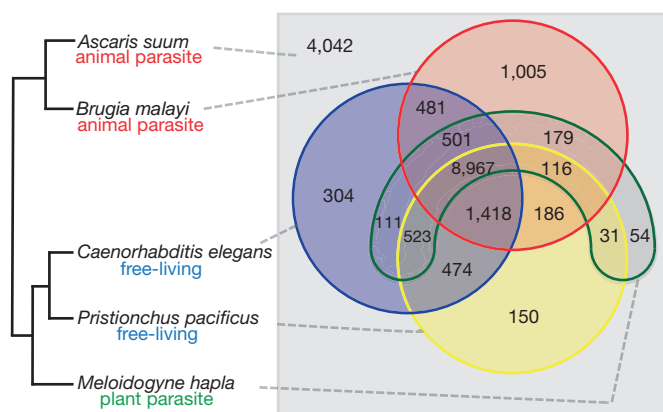
>2 kb) (Table 1) with a mean GC-content of 37.9%. This genome has few repetitive sequences (about 4.4% of the total assembly) relative to that reported for other metazoan genomes sequenced to date<sup>3–6</sup>, probably as a result of chromatin diminution<sup>7</sup>. We identified 424 distinct retrotransposon sequences (see Supplementary Tables 1–3) representing at least 22 families (8 long terminal repeats (LTRs), 12 long interspersed elements (LINEs) and 2 short interspersed elements (SINEs)), with *Gypsy*, *Pao* and *Copia* classes predominating for LTRs ( $n = 97$ , 85 and 60, respectively) and CR1, L1, and reverse transcriptase encoding RTE-RTE classes predominating for non-LTRs ( $n = 29$ , 28 and 21, respectively). We also identified eight families of DNA transposons (91 distinct sequences in total), of which *MuDr*, *En-Spm* and *Merlin* ( $n = 12$ , 9 and 8, respectively) predominated. We predicted 18,542 genes (14,783 supported by transcriptomic data), with a mean total length of 6.5 kb, exon length of 153 bp and a mean of 6.4 exons per gene (see Supplementary Fig. 2). Compared with the nematodes (roundworms) *Caenorhabditis elegans*<sup>3</sup>, *Pristionchus pacificus*<sup>8</sup>, *Brugia malayi*<sup>9</sup> or *Meloidogyne hapla*<sup>10</sup>, overall, the *A. suum* genes are significantly longer (see Supplementary Table 2), relating primarily to expansions of intronic regions (mean 1.1 kb).

Most (78.2%) of the predicted *A. suum* genes (Fig. 1) have a homologue (BLASTp cut-off  $\leq 10^{-5}$ ) either in *C. elegans* ( $n = 12,779$ ; 68.9%), *B. malayi* (12,853; 69.3%), *M. hapla* (10,482; 56.5%) or *P. pacificus* (11,865; 64.0%), with 8,967 being homologous among all species examined, and 4,042 (21.8%) being 'unique' to *A. suum* (see Fig. 1). Of the genes with homology to *C. elegans* or *B. malayi*, ~50%

**Table 1 | Features of the *Ascaris suum* draft genome**

Estimated genome size in megabases	309
Total number of base pairs within assembled scaffolds	272,782,664
N50 length in bp; total number >2 kb in length	407,899; 1,618
N90 length in bp; total number >N90 length	80,017; 748
GC content of whole genome (%)	37.9
Repetitive sequences (%)	4.4
Proportion of genome that is coding (exonic; including introns) (%)	5.9; 44.2
Number of putative coding genes	18,542
Gene size (mean bp)	6,536
Average coding domain length (mean bp)	983
Average exon number per gene (mean)	6
Gene exon length (mean bp)	153
Gene intron length (mean bp)	1,081
GC content in coding regions (%)	45
Number of transfer RNAs	255

N50 means 50% of all nucleotides in the assembly are within contigs of  $\geq 408$  kb. N90 means 90% of all nucleotides in the assembly are within contigs of  $\geq 80$  kb. Genome size estimated on the basis of  $k$ -mer (see online-only Methods) frequency.



**Figure 1 | Venn diagram summarizing the overlapping homology between the *Ascaris suum* gene set and those of other nematodes.** Grey box (right) represents genes unique to *A. suum*, relative to *Brugia malayi* (red circle), *Caenorhabditis elegans* (blue circle), *Meloidogyne hapla* (green arc) and/or *Pristionchus pacificus* (yellow circle). The phylogram (left) displays the evolutionary relationships currently proposed among the nematodes.

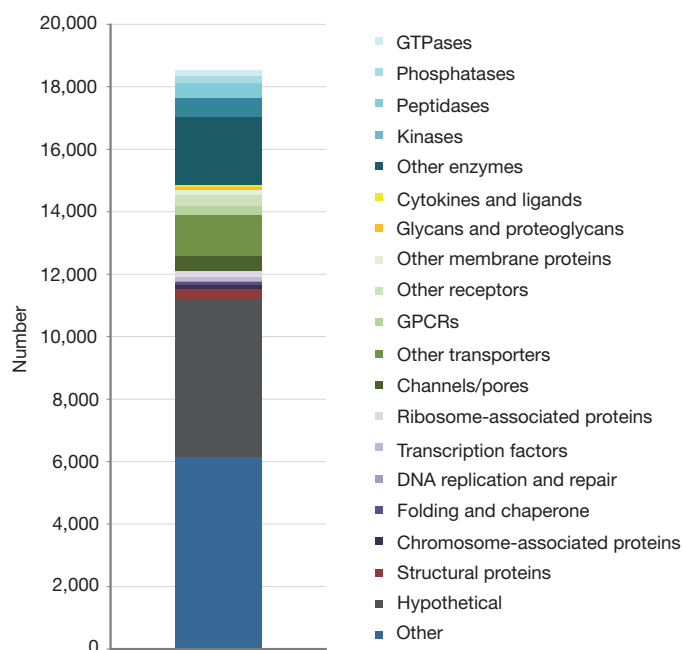
<sup>1</sup>Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria 3010, Australia. <sup>2</sup>BGI-Shenzhen, Shenzhen, 518083, China. <sup>3</sup>Ontario Institute for Cancer Research, MaRS Centre, South Tower, 101 College Street, Suite 800, Toronto, Ontario, M5G 0A3, Canada. <sup>4</sup>Laboratory of Parasitology and Parasitic Diseases, University of Ghent, Merelbeke B-9820, Belgium. <sup>5</sup>California Institute of Technology, Pasadena, California, 91125, USA. <sup>6</sup>Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. <sup>7</sup>Faculty of Life Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark.

\*These authors contributed equally to this work.



and 44%, respectively, were determined to represent one-to-one orthologues<sup>11</sup> (see Supplementary Data 1). For these orthologues (on scaffolds exceeding one megabase, 1 Mb, in size), we explored synteny for *A. suum* and *B. malayi* by pairwise comparison with *C. elegans* (see Supplementary Data 1). The findings show that interchromosomal gene rearrangements in *A. suum* are relatively rare and occurred less frequently in *A. suum* than in *B. malayi*<sup>9</sup> relative to *C. elegans* since their evolutionary divergence<sup>12</sup>. In contrast, intrachromosomal rearrangements were relatively common and comparable in frequency to those inferred for *B. malayi*<sup>9</sup>. Overall synteny was significantly higher between *A. suum* and *B. malayi* (~15%) than between either species and *C. elegans* (~3%), which is consistent with current knowledge of the evolutionary relationships among these three species<sup>12</sup>. Interestingly, of these *C. elegans* orthologous genes, 532 and 483 were exclusive to the current assemblies of the *A. suum* and *B. malayi* genomes, respectively (Supplementary Data 2). Although there were no homology matches between these two exclusive subsets of orthologues, they shared striking similarity in functional ontology (biological process), being linked predominantly to growth, reproduction, development and/or morphogenesis. There is clear evidence of plasticity in the germline of metazoans<sup>13</sup>, with cases of products from non-homologous genes in different species having analogous function(s). Therefore, we hypothesize that these two unique gene subsets relate to differences in reproductive biology (oviparity versus viviparity) and life history (direct versus indirect) between *A. suum* and *B. malayi*. Clearly, this proposal warrants testing and functional validation in *C. elegans* and/or in *Ascaris*.

Of the entire *A. suum* gene set, 2,370 genes had an orthologue (BLASTp cut-off  $\leq 10^{-5}$ ) belonging to one of 279 known biological (KEGG; see online-only Methods) pathways (Supplementary Data 3). Mapping to pathways in *C. elegans* indicated a full complement of molecules; by inference, the vast majority (95%) of the *A. suum* euchromatin is represented in the present genomic assembly, an inference that is supported by our transcriptomic data (Supplementary Tables 4 and 5). We were able to assign possible functions (such as for enzymes, receptors, channels and transporters; Supplementary Fig. 3, Supplementary Table 6 and Supplementary Data 4) to 13,503 (72.8%) of the genes predicted for *A. suum* (Fig. 2). For these genes, we predicted 456 peptidases belonging to five major classes (aspartic, cysteine, metallo-, serine and threonine), with the metallo- ( $n = 184$ ;



**Figure 2 | The major protein classes representing the *Ascaris suum* gene set.**

41.0%) and serine proteases ( $n = 132$ : 30.0%) predominating (Supplementary Data 4). Notably, the secreted peptidases (such as the M12 'astacins', the S9 and S33 serine proteases, and the C1 and C2 cysteine proteases) are abundantly represented, and have key roles in tissue invasion and degradation (for example, during migration and/or feeding) and/or immune evasion/modulation in many parasites<sup>14,15</sup>.

In addition, we identified 609 kinases and 257 phosphatases, respectively (Supplementary Data 4). All major classes of kinases are represented, with the tyrosine (TK:  $n = 94$ ), casein (CK1:  $n = 83$ ), CMGC ( $n = 67$ ) and CAMK ( $n = 54$ ) being most abundant in *A. suum*. The phosphatome includes 17 receptor and 68 conventional tyrosine, 64 serine/threonine and 39 dual-specificity phosphatases. On the basis of homology with molecules in *C. elegans*, 169 GTPases are encoded in the *A. suum* genome, including 135 small GTPases (Ras superfamily) representing the Rab ( $n = 36$ ), Ras ( $n = 35$ ; plus 8 Ras-like), Rho ( $n = 17$ ; plus 9 Rho-like) or Ran ( $n = 6$ ) subfamilies. Examples of these homologues include *eft-1*, *fzo-1*, *glo-1* and *rho-1*, which have essential roles in embryonic, larval and/or reproductive development (see [www.wormbase.org](http://www.wormbase.org)).

Given their key roles, many of these enzymes are proposed as targets for anti-parasitic compounds and/or vaccines<sup>16–18</sup>. Equally, the range of receptor and channel proteins identified here are interesting because many common anthelmintics bind such targets<sup>19</sup>. Here, we predicted 279 G protein-coupled receptors (GPCRs) for *A. suum* and 477 channel or pore proteins (Supplementary Data 4), including 272 voltage-gated and 98 ligand-gated ion channels. Many voltage-gated ion channels are known targets for nematocidal drugs, such as macrocyclic lactones (for example, ivermectin) and levamisole, and an aminoacetonitrile derivative, monepantel, is the most recent example of a highly effective nematocide that binds to a ligand-gated ion channel<sup>19</sup>. Importantly, in the *A. suum* gene set, we found a homologue (*acr-23*) of the *C. elegans* monepantel receptor<sup>19</sup>, suggesting that this drug may kill *A. suum*. In addition, we detected 462 transporters (for example, small molecule porter proteins), of which the major facilitator ( $n = 155$ ), cation symporter ( $n = 71$ ) and resistance-nodulation-cell division ( $n = 56$ ) superfamilies were most abundant (Supplementary Data 4).

Excretory/secretory (E/S) peptides are central to understanding parasite–host interactions. We predicted the secretome of *A. suum* to comprise 775 proteins with diverse functions (Supplementary Data 5). Notable among them are 68 secreted proteases, including 20 SC clan serine proteases (S9 and S33 families), 18 MA clan metallo-proteases (M10, M12 and M41 families) and 5 CA/CD clan cysteine proteases (C1 and C13 families); see <http://merops.sanger.ac.uk/> for clan definitions.

Secreted proteases have known roles in host-tissue degradation, required for feeding, tissue-penetration and/or larval migration for a range of helminths<sup>14</sup>, including *Ascaris*<sup>2</sup>. In addition, they are involved in inducing and modulating host immune responses against parasitic helminths<sup>15</sup>, which are often Th2-biased<sup>20</sup>. From the current understanding of these responses<sup>15</sup>, we compiled a comprehensive list of *A. suum* E/S proteins homologous to helminth-secreted peptides with important immunogenic or immunomodulatory roles in host animals (Supplementary Table 7 and Supplementary Data 6). Such homologues represent about half of the predicted *A. suum* secretome. Most abundant among them are O-linked glycosylated proteins ( $n = 300$ ), many of which are heavily targeted by immunoglobulin (Ig) M antibodies and bound by various pattern recognition receptors associated with host dendritic cells responsible for the induction of a Th2 immune response<sup>15</sup>.

Other members of the *A. suum* secretome are predicted to direct or evade immune responses. These peptides include a close homologue of the E/S-62 leucyl aminopeptidase of the filarioid nematode *Acanthocheilonema viteae*, which has been shown to inhibit B-cell, T-cell and mast cell proliferation/responses, promote an alternative activation of the host macrophages, through the inhibition of the Toll-like receptor signalling pathway, and induce a Th2 response through

the inhibition of IL-12p70 production by dendritic cells<sup>15</sup>. Additional, immunomodulatory molecules predicted for *A. suum* (BLASTp cut-off  $\leq 10^{-5}$ ) include homologues of another B-cell inhibitor (that is, the *B. malayi* cystatin CPI-2), several TGF- $\beta$  and macrophage initiation factor mimics, numerous neutrophil inhibitors, various oxidoreductases, and five close homologues of platelet anti-inflammatory factor  $\alpha$  (ref. 15). Some *A. suum* E/S peptides are predicted to be involved in immune evasion; for instance, some mask parasite antigens by mimicking host molecules (such as several C-type lectins with close homology to vertebrate macrophage mannose or CD23 (low affinity IgE receptors<sup>15</sup>).

Taken together, these data indicate that *A. suum* has a large arsenal of E/S proteins that are likely to be involved directly in manipulating, blocking and/or evading immune responses in the host. Understanding the immunomolecular interplay between *A. suum* and its host, early in infection, particularly during hepatopulmonary migration, should pave the way for designing prophylactic interventions, such as vaccination.

*Ascaris* larvae undertake an extensive migration through their host's body before they establish as adults in the small intestine. Following the ingestion of infective eggs and their gastric passage, third-stage larvae (L3s)<sup>21</sup> hatch from eggs in the gut and penetrate the intestinal wall; they then undergo, via the bloodstream, an arduous hepatopulmonary migration. The complexity of this migration coincides with important developmental changes in the nematode<sup>2</sup>. Clearly, this migration requires tightly regulated transcriptional changes in the parasite. We explored this aspect by characterizing the transcription profiles of infective L3s (from eggs), L3s from the liver or lungs of the host, and fourth-stage larvae (L4s) from the small intestine (Supplementary Fig. 4, Supplementary Data 7). Notable among genes enriched during larval migration are various secreted peptidases linked to tissue-penetration and degradation during feeding and/or migration<sup>14</sup>, including three C1/C2, five M1, eight M12, fourteen S9 and five S33 clan members. Considering the complex nature of larval migration, a key role for molecules associated with chemosensory pathways is highly likely. Such molecules have been studied extensively in *C. elegans*<sup>22</sup>, with numerous homologues being identified here in larval transcripts (Supplementary Data 7). With few exceptions, all of these homologues relate to olfactory chemosensation of volatile compounds (for example, alcohols, aldehydes or ketones), suggesting that the olfactory detection of molecular gradients is central to the navigation of *A. suum* larvae during migration. Lastly, considering the substantial host attack against migrating *Ascaris* larvae, E/S proteins probably play crucial roles in immune modulation and/or evasion during hepatopulmonary migration. Many such genes, including *Bm-alt-1*, *Bm-cpi-2* and *mif-4*, are highly transcribed in *A. suum* larvae (see Supplementary Data 7), particularly in migrating L3s.

Because of the large size of the adult nematode (10–15 cm), we were able to explore transcription in the musculature and reproductive tracts of adult male and female *A. suum* individuals as well (Supplementary Fig. 5 and Supplementary Data 8). Among the male-enriched transcripts is a range of genes associated specifically with sperm and/or spermatogenesis, including *fer-1*, *spe-4*, *spe-6*, *spe-9*,

*spe-10*, *spe-15* and *spe-41*, *alg-4* and *msp-57* (see www.wormbase.org). Notable among the female-enriched transcripts is a large variety of genes associated with oogenesis/egg-laying (such as *cat-1*, *unc-54*, *cbd-1* and *pqn-74*), vulval development (such as *noah-1*, *nhr-25*, *cog-1* and *pax-3*) and/or embryogenesis (such as *cam-1* and *unc-6*; see www.wormbase.org). Although the functions of these genes have been explored in *C. elegans* (primarily a hermaphroditic nematode), this detailed insight into the tissue-specific transcription for a dioecious nematode is a major advance.

Analyses of these RNA-seq data revealed 163,777 single nucleotide polymorphisms (SNPs) in coding regions of the *A. suum* genome; 61% of them were synonymous, 7% non-synonymous and <0.1% termination codons (Supplementary Data 9). Some of the most variable genes in *A. suum* encoded ribosomal proteins ( $n = 44$ ), translation initiation factor (*tif*) eIF-3 subunits 3 and 5, *tif* TFIIH subunit H2 and *tif* IF-2, galectin-4 and galectin-9, the latter two of which are probably linked to immune evasion<sup>15</sup> and may indicate that antigenic variation is among the many strategies apparent in *Ascaris* to combat the host immune response. Interestingly, the high nucleotide variability linked to the key elements of translation machinery did not relate to a bias in synonymous SNPs, suggesting that many mutations accumulate in particular 'hotspots' and/or are tolerated, but do not compromise either the structure or the function of this machinery. The least variable genes encoded various (druggable)<sup>16,17</sup> serine/threonine phosphatases ( $n = 17$ ) as well as numerous receptors, channels and transporters, for which there was an unusually strong bias towards synonymous SNPs, reinforcing their potential as intervention targets.

Given our present reliance on a small number of drugs (for example, piperazine, pyrantel, albendazole and mebendazole) for the treatment of ascariasis, their repeated or excessive use might lead to resistance in *Ascaris* populations to some or all of these compounds<sup>23</sup>. As few new anthelmintics (that is, aminoacetylnitriles<sup>19</sup> and cyclooctodepsipeptides<sup>24</sup>) have been discovered in the past two decades using traditional screening methods, an effective, alternative means of drug discovery is urgently needed<sup>23</sup>. Genome-guided drug target or drug discovery has major potential to complement conventional screening and re-purposing. The goal of genome-guided analysis is to identify genes or molecules whose inactivation by one or more drugs will selectively kill parasites but not harm their host.

Because most parasitic nematodes are difficult to produce or maintain outside of their host, or to subject to gene-specific silencing by RNAi<sup>23</sup> or morpholinos<sup>25,26</sup>, direct functional assessment of essentiality (that is, they are needed for the nematode's survival) is not yet practical. However, essentiality can be inferred from functional information for model organisms (for example, lethality in *C. elegans* and *D. melanogaster*)<sup>27</sup>, and this approach has indeed yielded effective targets for nematocides<sup>16</sup>. In *Ascaris*, we identified 629 proteins (Supplementary Data 10) with essential homologues in *C. elegans* and *D. melanogaster* (linked to lethal phenotypes upon gene perturbation). Among these are 87 channels or transporters (including 44 voltage-gated ion channels), which represent protein classes most successfully targeted for anthelmintic compounds, including macrocyclic lactones, levamisoles and aminoacetonitrile

**Table 2 | Druggable candidates in the *Ascaris suum* draft genome**

Protein or chokepoint	Subtype (number of molecules)	Total number
GTPase	Small GTPase (22); Ras (13); Rab (5); Rho (3); Ras-like (1); others (2)	46
Kinase	TK (8); AGC (3); CAMK (2); TKL (2); STE (1); other (1)	17
Peptidase	A22A (5); M14B (3); M12B (2); M67A (1); C14A (1); C50 (1); M12A (1); M13 (1); T01A (1); C46 (1); S33 (1)	19
Phosphatase	STP (28); cPTP (4); DSP (3)	35
Transporters and channels	Channels and pores (30); primary active transporters (24); incompletely characterized transport system proteins (22); accessory factors involved in transport (5); electrochemical potential-driven transporters (5); group translocators (1)	87
'Lethal' chokepoints	CDP-diacylglycerol-inositol 3-phosphatidyltransferase	1
	G protein-coupled receptor kinase 5	1
	Phosphoribosylformylglycinamide synthase	1
	Inosine-5'-monophosphate dehydrogenase	1
	Phospho-N-acetylmuramoyl pentapeptide transferase	1

Candidates were inferred from essentiality prediction and metabolic chokepoint analysis.

derivatives<sup>19,28</sup>. Also notable are 46 GTPases, 35 phosphatases (including PP1 and PP2A homologues, as targets for norcantharidin analogues)<sup>16</sup>, 17 kinases and 19 peptidases (Table 2).

In addition to essentiality-based prediction, an alternative strategy has been to infer enzymatic chokepoints intrinsic to the complete metabolome of a parasite<sup>29</sup>. Such chokepoints are defined as enzymatic reactions that uniquely produce and/or consume a molecular compound, using the strategy that the disruption of such enzymes would lead to the toxic build-up (that is, for unique substrates) or starvation (that is, for unique products) of metabolites within cells. Pathway analysis identified 225 likely chokepoints linked to genes predicted to be essential in *A. suum* (Supplementary Data 10). We gave the highest priority to targets predicted from single-copy genes in the *A. suum* genome, reasoning that lower allelic variability would exist within populations and would thus be less likely to give rise to drug resistance.

Using this strategy, we identified five high-priority drug targets for *A. suum* (see Table 2 and Supplementary Data 10) that, given their conservation with *C. elegans* and *D. melanogaster*, are likely to be relevant in relation to many other parasitic helminths. Conspicuous among them is IMP dehydrogenase (GMP reductase), which has a variety of inhibitors (for example, mycophenolic acid analogues<sup>30</sup>) that could be tested for ascaricidal effects. Clearly, the druggable genome of *Ascaris* now provides a solid basis for rational drug design, aimed at controlling parasitic nematodes of major socioeconomic impact worldwide.

In conclusion, we have characterized the genome of *A. suum*, a major parasite of one of the world's most important food animals (pig) and the closest relative of *A. lumbricoides*, which infects about 1.2 billion people globally<sup>1,2</sup>. Intriguingly, the present *A. suum* draft genome exhibits unusually low repeat content and lacks Tas2 transposons<sup>7</sup>. These characteristics probably relate to the chromatin diminution described previously for some ascaridoids<sup>7</sup>, indicating that our assembly represents the somatic genome of this parasite. The precise mechanism governing this diminution is not yet understood. Although the chromatin lost during this process is not fully characterized, there appears to be a significant loss in repeat content<sup>7</sup>, consistent with the present assembly. Notably, the present gene set inferred for *A. suum* includes *fert-1* and *rpS19G*, which, although originally proposed to be germline-specific<sup>7</sup>, were transcribed in all adult libraries sequenced here. This finding suggests that the genomic content lost during diminution might vary among individuals or tissues, and is a stimulus to investigate chromatin diminution between and among individual cells (that is, sperm or eggs), stages and tissue types of *A. suum*. Importantly, the present study, showing that a high-quality genomic assembly can be achieved using an approach based on whole-genome amplification, provides unique prospects for exploring diminution in detail, using the present genome as a reference.

In addition, our sequencing effort has characterized a broad range of key classes of molecules of major relevance to understanding the molecular biology of *A. suum* and the exquisite complexities of the host–parasite interplay on an immunobiological level. This work paves the way for future fundamental molecular explorations and the design of new methods for the treatment and control of one of the world's most important parasitic nematodes. This focus is now crucial, given the major impact of *Ascaris* and other soil-transmitted helminths, which affect billions of people and animals worldwide. Although these parasites are seriously neglected, genomic and post-genomic approaches provide new hope for the discovery of intervention strategies, with major implications for improving global health.

## METHODS SUMMARY

We sequenced the genome of *A. suum* using Illumina technology from genomic DNA from the reproductive tract of a single adult female. From six paired-end sequencing libraries (insert sizes: 0.17 kb to 10 kb; see Supplementary Tables 1 and 2), we generated 39 Gb of useable short-read sequence data, equating to ~80-fold coverage of the 273-Mb genome. We assembled the short reads, constructed

scaffolds in a step-by-step manner, and then closed intra-scaffold gaps<sup>5</sup>. Transposable elements, non-coding RNAs and the protein-coding gene set were inferred using a combination of predictive modelling and a homology-based approach. Orthology and synteny analyses were conducted using established methods<sup>9,11</sup>. We sequenced messenger RNA from infective L3s (from eggs), migrating L3s from the liver or lungs of the host, and L4s from the small intestine, as well as muscle and reproductive tissues from adult male and female *A. suum*, and used these data to aid gene predictions, define SNPs and explore key molecules associated with larval migration, reproduction and development. All proteins predicted from the gene set were annotated using databases for conserved protein domains, gene ontology annotations and model organisms (that is, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Mus musculus*). Essentiality and drug target predictions were conducted using established or in-house methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 June; accepted 12 September 2011.**

**Published online 26 October 2011.**

- Hotez, P. J., Fenwick, A., Savioli, L. & Molyneux, D. H. Rescuing the bottom billion through control of neglected tropical diseases. *Lancet* **373**, 1570–1575 (2009).
- Crompton, D. W. *Ascaris* and ascariasis. *Adv. Parasitol.* **48**, 285–375 (2001).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311–317 (2010).
- Mitrev, M. *et al.* The draft genome of the parasitic nematode *Trichinella spiralis*. *Nature Genet.* **43**, 228–235 (2011).
- Müller, F. & Tobler, H. Chromatin diminution in the parasitic nematodes *Ascaris suum* and *Parascaris univalens*. *Int. J. Parasitol.* **30**, 391–399 (2000).
- Dieterich, C. *et al.* The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nature Genet.* **40**, 1193–1198 (2008).
- Ghedini, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317**, 1756–1760 (2007).
- Opperman, C. H. *et al.* Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl Acad. Sci. USA* **105**, 14802–14807 (2008).
- Kuzniar, A., van Ham, R. C. H. J., Pongor, S. & Leunissen, J. A. M. The quest for orthologs: finding the corresponding gene across genomes. *Trends Genet.* **24**, 539–551 (2008).
- Blaxter, M. L. *et al.* A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**, 71–75 (1998).
- Ewen-Campen, B., Schwager, E. E. & Extavour, C. G. The molecular machinery of germ line specification. *Mol. Reprod. Dev.* **77**, 3–18 (2010).
- McKerrow, J. H., Caffrey, C., Kelly, B., Loke, P. & Sajid, M. Proteases in parasitic diseases. *Annu. Rev. Pathol.* **1**, 497–536 (2006).
- Hewitson, J. P., Grainger, J. R. & Maizels, R. M. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* **167**, 1–11 (2009).
- Campbell, B. E. *et al.* Norcantharidin analogues with nematocidal activity in *Haemonchus contortus*. *Bioorg. Med. Chem. Lett.* **21**, 3277–3281 (2011).
- Campbell, B. E., Hofmann, A., McCluskey, A. & Gasser, R. B. Serine/threonine phosphatases in socioeconomically important parasitic nematodes—prospects as novel drug targets? *Biotechnol. Adv.* **29**, 28–39 (2011).
- Renslo, A. R. & McKerrow, J. H. Drug discovery and development for neglected parasitic diseases. *Nature Chem. Biol.* **2**, 701–710 (2006).
- Kaminsky, R. *et al.* A new class of anthelmintics effective against drug-resistant nematodes. *Nature* **452**, 176–180 (2008).
- Maizels, R. M. & Yazdanbakhsh, M. Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nature Rev. Immunol.* **3**, 733–744 (2003).
- Geenen, P. L. *et al.* The morphogenesis of *Ascaris suum* to the infective third-stage larvae within the egg. *J. Parasitol.* **85**, 616–622 (1999).
- Bargmann, C. I. Chemosensation in *C. elegans* in *Wormbook* (ed. The *C. elegans* Research Community) (2006); <http://www.wormbook.org>.
- Keiser, J. & Utzinger, J. The drugs we have and the drugs we need against major helminth infections. *Adv. Parasitol.* **73**, 197–230 (2010).
- Harder, A. *et al.* Cyclooctadepsipeptides—an anthelmintically active class of compounds exhibiting a novel mode of action. *Int. J. Antimicrob. Agents* **22**, 318–331 (2003).
- Heasman, J. Morpholino oligos: making sense of antisense? *Dev. Biol.* **243**, 209–214 (2002).
- Geldhof, P. *et al.* RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects. *Parasitology* **134**, 609–619 (2007).
- Lee, I. *et al.* A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nature Genet.* **40**, 181–188 (2008).
- Campbell, W. C., Fisher, M. H., Stapley, E. O., Albers-Schonberg, G. & Jacob, T. A. Ivermectin: a potent new antiparasitic agent. *Science* **221**, 823–828 (1983).
- Berriman, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).



## METHODS

**Sample procurement, preparation and storage.** All specimens of *A. suum* were collected from pigs (*Sus scrofa*) with naturally acquired infections in Victoria, Australia (adult nematodes) and Ghent, Belgium (larval stages). L3s and L4s were also collected from the liver or lung and from the small intestine, respectively, of pigs, using established procedures<sup>31,32</sup>. Nematodes were washed extensively in sterile physiological saline (37 °C), snap-frozen in liquid nitrogen and then stored at -70 °C until use.

**DNA isolation, sequencing and quality control.** Total genomic DNA was isolated from the reproductive tract of a single adult female of *A. suum* using a sodium-dodecyl sulphate/proteinase K digestion<sup>33</sup> followed by phenol-chloroform extraction and ethanol precipitation<sup>34</sup>. Total DNA yield was determined using the Qubit fluorometer double-stranded DNA HS Kit (Invitrogen). DNA integrity was verified with a 2100 Bioanalyser (Agilent). Short-insert (170 bp and 500 bp) and mate-pair (800 bp, 2 kb, 5 kb and 10 kb) genomic DNA libraries were prepared and paired-end sequenced using TruSeq chemistry on a HiSeq 2000 (Illumina). Whole-genome amplification, employing the REPLI-g Midi Kit (Qiagen), was used to produce (from 200 ng of genomic template) the required amount of DNA for the construction of the 2-kb, 5-kb and 10-kb libraries (Supplementary Fig. 6). The sequence data generated from each of the six libraries were verified, and low-quality sequences, base-calling duplicates and adapters removed. The size of the genome and the heterozygosity rate were estimated by establishing the frequency of occurrence of each 17-bp *k*-mer (a unique sequence of *k* (that is, 17) nucleotides in length) within the genomic sequence data set (from the 170-bp library) using an established method<sup>5</sup>. Genome size was estimated using a modification of the Lander–Waterman algorithm<sup>35</sup>, where the haploid genome length in base pairs is  $G = (N \times (L - K + 1) - B)/D$ , where *N* is the read length sequenced in base pairs, *L* is the mean length of sequence reads, *K* is the *k*-mer length (17 bp) and *B* is the number of *k*-mers occurring less than four times (Supplementary Fig. 7). Heterozygosity was evaluated throughout the genome assembly by assessing the distribution of the *k*-mer frequency in the sequence data set.

**RNA isolation, sequencing and assembly.** We obtained total RNAs from egg-L3s (*n* ≈ 500,000), liver-L3s (*n* ≈ 60,000), lung-L3s (*n* ≈ 80,000) or L4s (*n* ≈ 30,000) and from the somatic musculature or reproductive tract of each of two adult male and two adult female *A. suum* using the TriPure reagent (Roche), and both yield and quality were verified by 2100 BioAnalyser (Agilent). Polyadenylated (polyA+) RNA was purified from 10 µg of total RNA using Sera-mag oligo(dT) beads, fragmented to a size of 300–500 bp, reverse-transcribed using random hexamers, end-repaired and adaptor-ligated, according to the manufacturer's protocol (Illumina). Ligated products of ~400 bp were excised from agarose and then PCR-amplified (15 cycles), as recommended. Products were purified over a MinElute column (Qiagen) and subjected to paired-end RNA-seq using TruSeq chemistry on a HiSeq 2000 (Illumina) and assessed for quality and adaptor sequence. Transcripts were assembled from RNA-seq data using Oases<sup>36</sup>. All transcripts were used to assess the completeness of the genome assembly and to predict genes.

**Genomic assembly and quality control.** Following sequencing, all DNA-sequence reads were corrected based on *k*-mer (=17) distribution<sup>5</sup>. Briefly, sequence reads were removed if >10% of bases were ambiguous (represented by the letter N) or multiple adenosine monophosphates (poly-A), and all remaining reads were filtered on the basis of Phred quality. For small insert-size libraries (that is, <800 bp), additional reads were removed from the final data set if >65% of bases were of a low Phred quality (<8). For large insert libraries (2 kb, 5 kb and 10 kb), reads were removed from the final data set if >80% of bases were of a low Phred quality (<8). Duplicate (that is, identical) reads and partial reads representing the Illumina adaptor sequence were also removed, as were reads from the 500-bp library representing paired reads found to overlap by >10 bp (allowing for a 10% mismatch). Corrected and filtered data were assembled into contigs using SOAPdenovo<sup>3</sup>, and joined iteratively into scaffolds using a step-wise process (see Supplementary Fig. 8), using the paired reads generated from each library; local assemblies were used to close all gaps. Each nucleotide position in the final assembly was assessed for accuracy by aligning all filtered reads to the scaffolds using SOAP2aligner<sup>37</sup>, allowing for up to five mismatches per read. The depth of coverage and repeat content were assessed initially by sliding-window analysis and presented as a frequency distribution (Supplementary Fig. 9). GC-content was estimated using 10-kb non-overlapping sliding windows, and GC-bias<sup>38</sup> was assessed based on a frequency distribution of these data (Supplementary Fig. 10). To assess the completeness of the genome assembly, RNA-seq data representing each of the organs (that is, musculature and reproductive tract), genders and/or stages of *A. suum* sequenced were mapped to the final assembly using the BLAST-like Alignment Tool (BLAT)<sup>39</sup>.

**Assessment of repeat content and annotation of non-coding RNA.** Following genome assembly, tandem repeats were identified using the Tandem Repeats Finder program<sup>40</sup>. Transposable elements were predicted using a combination of homology-based comparisons (using RepeatMasker<sup>41</sup>) and *de novo* approaches (using LTR\_FINDER<sup>42</sup>, PILER<sup>43</sup> and RepeatScout<sup>44</sup>), with a consensus population of predicted repetitive elements, constructed in RepeatScout using fit-preferred alignment scores. Low-frequency repeats (≤25) and multi-copy genes (in the repeat element library) were filtered using RepeatMasker, producing a non-redundant sequence file, which was then used to identify and classify additional homologous repeats in the genome.

**Gene prediction, and syntenic and genetic variation analysis.** The *A. suum* protein-coding gene set was inferred using *de novo*-, homology- and evidence-based (that is, transcriptomic) approaches. *De novo* gene prediction was performed on a repeat-masked genome using three programs (Augustus, GlimmerHMM and SNAP)<sup>5</sup>; training models were generated from a subset of the transcriptomic data set representing 1,355 distinct genes. Homology-based prediction was conducted by comparison with complete genomic data for *Caenorhabditis elegans*<sup>3</sup>, *Pristionchus pacificus*<sup>8</sup> and *Brugia malayi*<sup>9</sup> using a multi-phase strategy, in which (1) all putative homologous gene sequences were preliminarily identified from alignments with protein sequences representing the complete gene set of each of the reference genomes (the longest transcripts were chosen to represent each gene) by TblastN (*e*-value cut-off: 10<sup>-5</sup>) and grouped into gene-like structures using genBlastA<sup>45</sup>; (2) regions representing these putative genes, and flanking regions (3,000 bp) at the 5'- and 3'-ends of each predicted gene, were extracted from the assembly and aligned to the 'parent' sequences derived from the reference genomes using Genewise<sup>46</sup>; (3) all single-exon genes predicted to have arisen from a retro-transposition and containing at least one frame-shift error or representing incomplete coding domains of <150 bp as well as all multi-exon genes containing more than two frame-shift errors and/or representing incomplete coding domains of <100 bp, were discarded. Evidence-based gene prediction was conducted by aligning all RNA-seq data generated herein against the assembled genome using TopHat<sup>47</sup>, with cDNAs predicted from the resultant data using Cufflinks<sup>48</sup>. Following the prediction of genes, a non-redundant gene set representing homology-based, *de novo*-predicted and RNA-seq-supported genes, was generated using Glean (<http://sourceforge.net/projects/glean-gene/>)<sup>5</sup>. All Glean-predicted genes were retained, as were all genes supported by RNA-seq data and those predicted using two or more *de novo* methods (that is, Augustus, GlimmerHMM and/or SNAP). The open reading frame of each gene was predicted using BestORF ([www.softberry.com](http://www.softberry.com)). To assess the quality and accuracy of the predicted gene set, we examined the length-distribution of all genes, coding sequences, exons and introns, and the distribution of exon numbers for individual genes, and then compared these parameters with those calculated for the published gene sets of *B. malayi*, *C. elegans*, *P. pristionchus* and *M. incognita* (Supplementary Fig. 4).

Following prediction of the finalized gene set, we conducted pairwise analysis of the overall syntenic existing between/among the large (>1 Mb) assembly scaffolds for *B. malayi* and *A. suum* relative to the complete *C. elegans* chromosomes. This analysis was undertaken by conducting pairwise alignments among all *A. suum* or *B. malayi* (WS220 assembly: [ftp://ftp.sanger.ac.uk/pub2/wormbase/releases/WS220/genomes/b\\_malayi/](ftp://ftp.sanger.ac.uk/pub2/wormbase/releases/WS220/genomes/b_malayi/)) scaffolds larger than 1 Mb in size and the *C. elegans* chromosomes using LASTz ([http://www.bx.psu.edu/miller\\_lab/dist/README.lastz-1.02.00/README.lastz-1.02.00a.html](http://www.bx.psu.edu/miller_lab/dist/README.lastz-1.02.00/README.lastz-1.02.00a.html)), which were then joined using CHAINNET<sup>49</sup> and output as a .axt alignment from which large-syntenic regions were defined. The resulting alignment files were used to construct syntenic images on scaleable vector graphics format using customized perl scripts (ZX). In addition, gene-level syntenic analyses were conducted for one-to-one orthologous genes colocalizing to large *A. suum* or *B. malayi* assembly scaffolds (>1 Mb) according to ref. 9. Orthology was determined by pairwise reciprocal BLASTx comparisons between *A. suum*, *B. malayi* and *C. elegans* according to ref. 11. One-to-one orthologous genes shared between either *A. suum* or *B. malayi* and *C. elegans* but not shared among *A. suum* and *B. malayi* based on reciprocal BLASTp analysis were further confirmed by Hidden Markov Modelling using the jackhmmer command in the program HMMER 3.0 (ref. 50) and a highly permissive threshold (HMM cutoff: 10<sup>-2</sup>).

We assessed the genome-wide variation in the exonic regions by mapping all raw reads from our transcriptomic data to the genomic coding domains using Maq<sup>51</sup>, and calling SNPs with a minimum coverage threshold of ten reads. All mapped reads were assessed as synonymous (non-coding change), non-synonymous (coding change) or ambiguous (a SNP that was represented in our data set as an ambiguous IUPAC code wherein one nucleotide change would cause a synonymous mutation and the other a non-synonymous mutation) using a custom Perl script (snp\_analysis.pl). All genes were then ranked based on their accumulation of SNPs to assess and identify their levels of conservation/variation relative to their

function. We reasoned that, in addition to the real effects of the variability of each gene on their accumulation of SNPs, these data would be influenced also by the coverage achieved for each gene, which is affected by the number of reads available for each gene (that is, their relative levels of transcription) and the length of each gene. Thus, before ranking, the SNP data for each gene was normalized for its calculated reads per kilobase per million reads (RPKM) and total gene length using the simple equation: SNPs per read per kilobase = total SNPs divided by RPKM divided by gene length (in bp) multiplied by 1,000 bp. Following ranking, we explored function among the 2.5% most variable (with the highest rankings based on normalized SNP data) and most conserved genes (with the lowest rankings based on normalized SNP data). Noting the potential inaccuracy associated with estimating the normalized SNP rankings of lowly transcribed genes (owing to a lack of data/coverage), only genes for which at least 100 reads were available were considered in these functional comparisons.

**Functional annotation of coding genes.** Following the prediction of the protein-coding gene set, each inferred amino acid sequence was assessed for conserved protein domains in the SPOT, Pfam, PRINTS, PROSITE, ProDom and SMART databases using InterProScan<sup>52</sup>, employing default settings. Gene ontology categories<sup>53</sup> were assigned to each contig inferred to contain at least one conserved protein domain. Gene ontology categories were summarized and standardized to level 2 and level 3 terms, defined using the GOSlim hierarchy<sup>54</sup> using WEGO<sup>55</sup>. To characterize further the contigs/transcripts from *A. suum*, we conducted a series of high-stringency BLASTp homology searches (*e*-value cut-off:  $10^{-5}$ ) against a variety of databases. Each contig was assessed for a known functional orthologue, defined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.kegg.com). Where appropriate, orthologous matches were mapped visually to a defined pathway using the KEGG pathway tool (available via www.kegg.com) or clustered to a known protein family using the KEGG-BRITE hierarchy tool (available via www.kegg.com). In addition, the amino acid sequence inferred from each *A. suum* coding gene was compared by BLASTp with protein sequences available for key nematode species (*B. malayi*, *C. elegans*, *P. pacificus* and *M. incognita*) as well as for *Drosophila melanogaster*<sup>56</sup> and *Mus musculus*<sup>56</sup> and those contained within the UniProt<sup>57</sup>, SwissProt and TrEMBL databases<sup>58</sup>. Key protein groups (for example, peptidases, kinases, phosphatases, GTPases, GPCRs, and transport and channel proteins) were characterized by high-stringency BLASTp homology searching (*e*-value cut-off  $<10^{-5}$ ) of manually curated information sequence data available in the MEROPS<sup>59</sup>, WormBase, KS-Sarfari (https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari) and GPCR-Sarfari (https://www.ebi.ac.uk/chembl/sarfari/gpcrsarfari) and the Transporter Classification database<sup>60</sup>. E/S proteins were predicted using Phobius<sup>61</sup>, employing both the neural network and hidden Markov models, and by BLASTp homology-searching of the validated signal peptide database<sup>62</sup> and an E/S database containing published proteomic data for *B. malayi*<sup>63,64</sup>, *Schistosoma mansoni*<sup>65</sup> and *M. incognita*<sup>66</sup>. In the final annotation, proteins inferred from genes were classified based on a homology match (*e*-value cut-off:  $\leq 10^{-5}$ ) to: (1) a curated, specialist protein database, followed by (2) the KEGG database, followed by (3) the UniProt/SwissProt/TrEMBL databases, followed by (4) the annotated gene set for a model organism, including *C. elegans*, *D. melanogaster*, *M. musculus* or *S. cerevisiae*, followed by (5) the gene ontology classification, and, finally, (6) a recognized, conserved protein domain based on InterProScan analysis. Any inferred proteins lacking a match (BLASTp cut-off  $\leq 10^{-5}$ ) in at least one of these analyses were designated hypothetical proteins. The final annotated protein-coding gene set for *A. suum* is available for download at WormBase (in nucleotide and amino acid formats).

**Differential transcription analysis.** Following RNA-seq, all paired-end reads for each library constructed were aligned to the predicted *A. suum* gene set using TopHat, and quantitative levels of transcription (RPKM)<sup>67</sup> were calculated using Cufflinks. Differential transcription was assessed<sup>68</sup> using a *P*-value cut-off of  $\leq 0.01$  and a minimum, two-fold difference in absolute RPKM values. False discovery rates for differential transcription were determined<sup>68</sup>. To allow the rapid visual assessment of the statistically significant changes in transcription of each gene between and among individual libraries, we constructed heat-maps representing absolute differences in the RPKM values, calculated for each transcript using a customized Perl script (express\_heatmap\_RPKM.pl). Genetic interaction networks were predicted<sup>69</sup> based on data available for homologous genes in *C. elegans* (inferred from BLASTp comparisons) and viewed using the program BioLayout 3D<sup>70</sup>.

**Essentiality and druggability predictions.** *A. suum* genes with homology to those in the *C. elegans* and/or *D. melanogaster* genomes were inferred based on BLASTp comparisons using the predicted protein sequences for individual species (*e*-value cut-off  $10^{-5}$ ). Phenotypic data for each *C. elegans* and *D. melanogaster* homologue were sourced from WormBase and FlyBase (www.flybase.org), respectively. *A. suum* genes determined<sup>71</sup> to have homologues with lethal phenotypes in both *C. elegans* and *D. melanogaster* were inferred to represent essential

genes. Metabolic chokepoints were defined<sup>29,72</sup> and assessed based on *A. suum* gene sequences determined, by BLASTp comparison ( $10^{-5}$ ), to have an orthologue in the KEGG database. All 'essential' homologues and/or molecules in 'choke-points' were then queried against the BRENDA<sup>73</sup> and ChEMBL databases (accessible via https://www.ebi.ac.uk/chembl/bd/), to identify known chemical inhibitors.

**Additional bioinformatic analyses, and use of software.** Data analysis was conducted in a Unix environment or Microsoft Excel 2007 using standard commands. Bioinformatic scripts required to facilitate data analysis were designed using Perl, BioPerl, Java and Python and are available via http://research.vet.unimelb.edu.au/gasserlab/.

- Cantacessi, C. *et al.* Differences in transcription between free-living and CO<sub>2</sub>-activated third-stage larvae of *Haemonchus contortus*. *BMC Genom.* **11**, 266, doi:10.1186/1471-2164-11-266 (2010).
- Saeed, I., Roepstorff, A., Rasmussen, T., Hog, M. & Jungersen, G. Optimization of the agar-gel method for isolation of migrating *Ascaris suum* larvae from the liver and lungs of pigs. *Acta Vet. Scand.* **42**, 279–286 (2001).
- Gasser, R. B. *et al.* Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature Protocols* **1**, 3121–3128 (2007).
- Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* 3rd edn, Vol. 3, E.3–E.4 (Cold Spring Harbor Laboratory, 2001).
- Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
- Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
- Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
- Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protocols Bioinformatics* Ch. 4.10 (2009).
- Xu, Z. & Wang, H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
- Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21** (Suppl. 1), i152–i158 (2005).
- Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21** (Suppl. 1), i351–i358 (2005).
- She, R., Chu, J. S., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
- Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
- Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
- Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100**, 11484–11489 (2003).
- Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **23**, 205–211 (2009).
- Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
- Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
- Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
- Camon, E. *et al.* The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res.* **13**, 662–672 (2003).
- Ye, J. *et al.* WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297 (2006).
- Chinwalla, A. T. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Wu, C. H. *et al.* The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* **34**, D187–D191 (2006).
- Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
- Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227–D233 (2010).
- Saier, M. H. Jr, Yen, M. R., Noto, K., Tamang, D. G. & Elkan, C. The Transporter Classification Database: recent advances. *Nucleic Acids Res.* **37**, D274–D278 (2009).
- Kall, L., Krogh, A. & Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).
- Chen, Y. *et al.* SPD—a web-based secreted protein database. *Nucleic Acids Res.* **33**, D169–D173 (2005).
- Bennuru, S. *et al.* *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl. Trop. Dis.* **3**, e410 (2009).



# Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie<sup>1\*</sup>, Mark W. Barnett<sup>1\*</sup>, Kyle R. Upton<sup>1\*</sup>, Daniel J. Gerhardt<sup>2</sup>, Todd A. Richmond<sup>2</sup>, Fioravante De Sapia<sup>1</sup>, Paul Brennan<sup>3</sup>, Patrizia Rizzu<sup>4</sup>, Sarah Smith<sup>1</sup>, Mark Fell<sup>1</sup>, Richard T. Talbot<sup>1</sup>, Stefano Gustincich<sup>5</sup>, Thomas C. Freeman<sup>1</sup>, John S. Mattick<sup>6</sup>, David A. Hume<sup>1</sup>, Peter Heutink<sup>4</sup>, Piero Carninci<sup>7</sup>, Jeffrey A. Jeddell<sup>2</sup> & Geoffrey J. Faulkner<sup>1</sup>

**Retrotransposons are mobile genetic elements that use a germline ‘copy-and-paste’ mechanism to spread throughout metazoan genomes<sup>1</sup>. At least 50 per cent of the human genome is derived from retrotransposons, with three active families (L1, *Alu* and SVA) associated with insertional mutagenesis and disease<sup>2,3</sup>. Epigenetic and post-transcriptional suppression block retrotransposition in somatic cells<sup>4,5</sup>, excluding early embryo development and some malignancies<sup>6,7</sup>. Recent reports of L1 expression<sup>8,9</sup> and copy number variation<sup>10,11</sup> in the human brain suggest that L1 mobilization may also occur during later development. However, the corresponding integration sites have not been mapped. Here we apply a high-throughput method to identify numerous L1, *Alu* and SVA germline mutations, as well as 7,743 putative somatic L1 insertions, in the hippocampus and caudate nucleus of three individuals. Surprisingly, we also found 13,692 somatic *Alu* insertions and 1,350 SVA insertions. Our results demonstrate that retrotransposons mobilize to protein-coding genes differentially expressed and active in the brain. Thus, somatic genome mosaicism driven by retrotransposition may reshape the genetic circuitry that underpins normal and abnormal neurobiological processes.**

Malignancy and ageing are commonly associated with the accumulation of deleterious mutations that lead to loss of function, cell death or uncontrolled growth. Retrotransposition is strongly mutagenic; an estimated 400 million retrotransposon-derived structural variants are present in the global human population<sup>3</sup> and more than 70 diseases involve heritable and *de novo* retrotransposition events<sup>2</sup>. Presumably for this reason, transposition-competent retrotransposons are heavily methylated and transcriptionally inactivated<sup>4,5</sup>. Nevertheless, substantial somatic L1 retrotransposition has been detected in neural cell lineages<sup>10–12</sup>. Given the complex structural and functional organization of the mammalian brain, its adaptive and regenerative capabilities<sup>13</sup> and the unresolved aetiology of many neurobiological disorders, these somatic insertions could be of great importance<sup>14</sup>.

One explanation for the observed transpositional activity in the brain may be that the L1 promoter is transiently released from epigenetic suppression during neurogenesis<sup>11,12</sup>. Transposition-competent L1 retrotransposons can then repeatedly mobilize to different loci in individual cells and produce somatic mosaicism. Several lines of evidence support this model, including L1 transcription<sup>8,9</sup> and copy number variation (CNV) in brain tissues from human donors of various ages<sup>10,11</sup>, as well as mobilization of engineered L1 retrotransposons *in vitro* and in transgenic rodents<sup>10,12</sup>. Importantly, it is not known where somatic L1 insertions occur in the genome, nor, considering that open chromatin is susceptible to L1 integration<sup>15</sup>, whether these events disproportionately affect protein-coding loci expressed in the brain.

Mapping the individual retrotransposition events that collectively form a somatic mosaic is challenging owing to the rarity of each mutant allele in a heterogeneous cell population. We therefore developed a high-throughput protocol that we call retrotransposon capture sequencing (RC-seq). First, fragmented genomic DNA was hybridized to custom sequence capture arrays targeting the 5' and 3' termini of full-length L1, *Alu* and SVA retrotransposons (Fig. 1a and Supplementary Tables 1 and 2). Immobile ERVK and ERV1 long terminal repeat (LTR) elements were included as negative controls. Second, the captured DNA was deeply sequenced, yielding ~25 million paired-end 101-mer reads per sample (Fig. 1b). Last, read pairs were mapped using a conservative computational pipeline designed to identify known (Fig. 1c) and novel (Fig. 1d and Supplementary Fig. 1a–d) retrotransposon insertions with uniquely mapped read pairs ('diagnostic reads') spanning their termini.

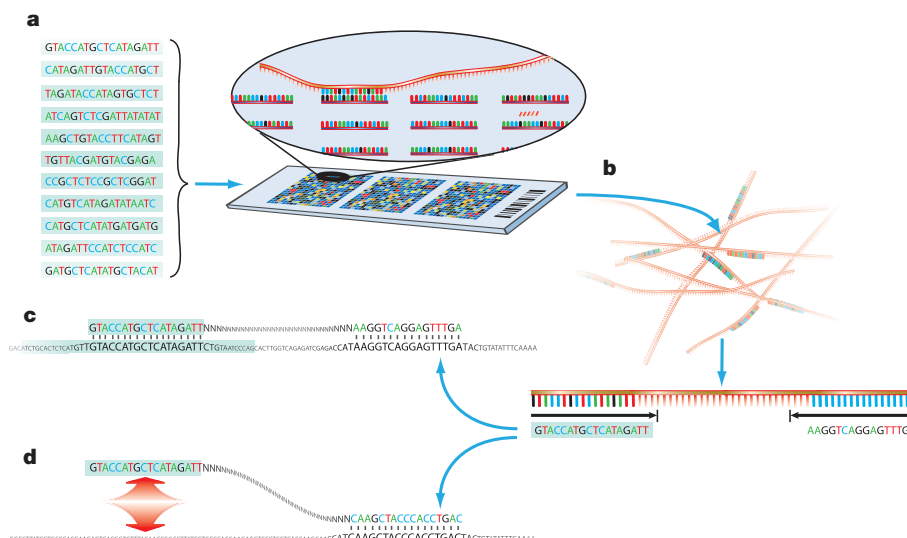
Previous works have equated L1 CNV with somatic mobilization *in vivo*<sup>10,11</sup>. To test this assumption with RC-seq, we first screened five brain subregions taken from three individuals (donors A, B and C) for L1 CNV. A significant ( $P < 0.001$ ) increase was observed in the number of copies of L1 open reading frame 2 (ORF2) present in DNA extracted from the hippocampus of donor C, and a similar, though smaller, increase was observed for donor A (Fig. 2). We then applied RC-seq to the brain regions that showed the highest (hippocampus) and lowest (caudate nucleus) L1 CNV using samples from all three donors, including a technical replicate of caudate nucleus from donor A. A total of 177.4 million RC-seq paired-end reads were generated from seven libraries (Supplementary Table 3). RC-seq achieved deep sequencing coverage of known active retrotransposons, high reproducibility and limited sequence capture bias (Supplementary Results).

Read pairs diagnostic for novel retrotransposon insertions were clustered on the basis of their insertion site, relative orientation and retrotransposon family. A total of 25,229 clusters were produced. Proximal clusters arranged on opposing strands indicated two termini of one insertion and were paired, resulting in a catalogue of 24,540 novel insertions (Supplementary Table 4). As expected, the great majority of these were either L1 (32.2%) or *Alu* (60.9%) (Fig. 3a). To segregate germline mutations from other events, we combined the three largest available catalogues of L1 and *Alu* polymorphisms<sup>6,16,17</sup> as an annotation database and also performed RC-seq on pooled genomic DNA extracted from blood, producing 6,150 clusters (Supplementary Table 5) that were intersected with the existing brain RC-seq clusters. Any brain clusters that contained RC-seq reads from more than one region or individual, overlapped a blood RC-seq cluster or matched a known polymorphism were designated as germline insertions. Overall, 8.4% of *Alu* insertions in the brain were annotated as germline, versus only 1.9% for L1. Nearly all unannotated

<sup>1</sup>Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK. <sup>2</sup>Roche NimbleGen, Inc., 500 South Rosa Road, Madison, Wisconsin 53719, USA. <sup>3</sup>Edinburgh Cancer Research Centre, Western General Hospital, Crewe Road South, Edinburgh EH4 2XR, UK. <sup>4</sup>Section of Medical Genomics, Department of Clinical Genetics, VU University Medical Center, Van der Boerhorststraat 7, 1081 BT Amsterdam, The Netherlands. <sup>5</sup>Sector of Neurobiology, International School for Advanced Studies (SISSA), via Bonomea 265, 34136 Trieste, Italy. <sup>6</sup>Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. <sup>7</sup>RIKEN Yokohama Institute, Omics Science Center, 1-7-22 Suehiro-chô, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.

\*These authors contributed equally to this work.





**Figure 1 | Overall RC-seq methodology.** **a**, Retrotransposon capture: sheared genomic DNA is hybridized to custom tiling arrays probing full-length retrotransposons (nucleotides highlighted with light-blue background). **b**, Sequencing: after hybridization, DNA fragments are eluted and analysed with an Illumina sequencer, producing  $\sim 2.5 \times 10^7$  paired-end reads per library

L1 insertions matched fewer than three diagnostic RC-seq reads (Fig. 3b) and were considered potential somatic insertions.

Candidate insertions were validated by PCR amplification and capillary sequencing. Thirty-five germline L1, *Alu*, SVA and LTR insertions were readily confirmed by single-step PCR (Supplementary Table 6). Given low target molecule abundance and the high genomic frequency of the L1 3' end, we devised a 5'-end nested PCR validation assay for somatic insertions. From 850 and 2,601 full-length ( $\geq 90\%$ ) L1 and *Alu* insertions, respectively representing 11.0% and 19.0% of the putative somatic insertions found for each family, we selected 29 examples (14 L1 and 15 *Alu*) for validation. Nearly all of the chosen examples were exonic or intronic and were prioritized on the basis of the degree of 5' truncation, with longer insertions preferred. Optimization of the protocol, combined with substantial input DNA (100 ng), ultimately led to the confirmation of all of the L1 insertions and 12 of the 15 *Alu* insertions (Supplementary Table 7 and Supplementary Fig. 2). Four somatic SVA insertions were also assayed using the same process, and two

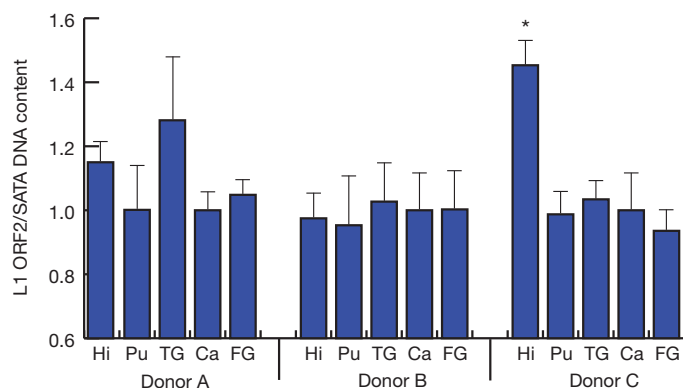
that are subsequently aligned to the reference genome. **c**, Reads mapping as a pair to a single locus indicate known retrotransposon insertions. **d**, Unpaired reads where one end maps to a single locus and the other end maps to a distal retrotransposon indicate novel retrotransposition events.

were confirmed (both from subfamily SVA\_F) before the available input material was exhausted.

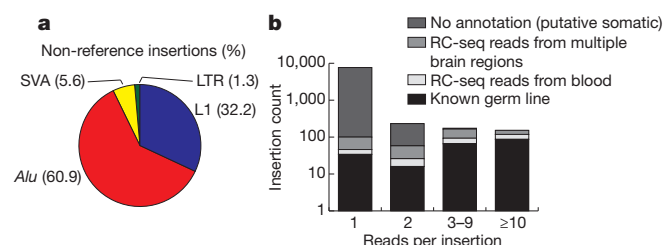
Repeated attempts to PCR-amplify the corresponding 3' junctions consistently yielded off-target amplicons, leaving validation based exclusively on 5' junctions. For this reason, we could not experimentally identify the target-site duplications (TSDs) that are indicative of retrotransposition through target-primed reverse transcription<sup>1</sup>. We propose that the 3' junctions of insertions validated at their 5' ends did not amplify efficiently owing to the confounding factors listed above, and to the presence of long polyA tails in on-target amplicons but often not, as we found, in off-target amplicons.

However, TSDs could in some cases be found directly by RC-seq (Supplementary Fig. 1d). An examination of germline insertions sequenced to high depth ( $\geq 10$  reads) at both their 5' and their 3' ends revealed that 43 of 50 (86%) presented TSDs. Owing to their very low abundance—and, therefore, low sequencing coverage—only three putative somatic insertions were detected by at least one RC-seq read at both termini. Two of these insertions (one L1 and one *Alu*) presented TSDs. Despite these and other data strongly supporting retrotransposition as the main cause of somatic mobilization (Supplementary Results), an insufficient number of insertions were sequenced at both ends to determine whether target-primed reverse transcription or an alternative retrotransposition mechanism<sup>18</sup> was primarily responsible.

The somatic origin of each insertion was demonstrated by its presence in one of the assayed brain tissues and its absence from the other, according to RC-seq and PCR results. As illustrative examples, an



**Figure 2 | Multiplex quantitative PCR confirms L1 CNV in the human brain.** The relative abundance of L1 open reading frame 2 (ORF2) with respect to  $\alpha$ -satellite repeats (SATA) was quantified using an existing TaqMan-based approach<sup>10</sup>. Genomic DNA from five brain regions was assayed in three donors (A, B and C). Hi, hippocampus; Pu, putamen; TG, middle temporal gyrus; Ca, caudate nucleus; FG, middle frontal gyrus. Values are normalized to caudate nucleus for each donor. Error bars, s.e.m. \* $P < 0.001$  for repeated-measures one-way analysis of variance within each donor, followed by pairwise least-significant-difference *post hoc* tests with Bonferroni correction.



**Figure 3 | Characterization of non-reference genome insertions.** **a**, Proportions of novel insertions found for each family. **b**, Annotation of novel L1 insertions (note logarithmic scale) across all brain libraries. The great majority of insertions detected by fewer than three reads could not be annotated and were considered putative somatic events.

intronic somatic L1 insertion in *HDAC1* is detailed in Fig. 4a, b and an exonic somatic *Alu* insertion in *RAI1* is shown in Fig. 4c, d. These experimental results indicated that insertions detected by RC-seq occurred *in vivo* and did not represent sequencing artefacts.

Donor element annotation revealed that 80.2% of somatic L1 insertions corresponded to the most recently active human L1 subfamilies, L1-Ta and pre-Ta (Supplementary Fig. 3a). The normalized hippocampus/caudate nucleus ratios for somatic L1 insertions were 1.3, 0.5 and 2.2 for donors A, B and C, respectively, paralleling trends from the L1 CNV assay (Fig. 2). Protein-coding loci were disproportionately affected (Supplementary Table 8) relative to random expectation and prior germline frequencies ( $P < 0.0001$  for exons and introns,  $\chi^2$ -test). Pre-existing microarray expression data indicated that genes containing intronic L1 insertions were twice as likely to be differentially overexpressed in the brain as would be randomly expected ( $P < 0.0001$ ,  $\chi^2$ -test). Key loci were found to contain somatic L1 insertions, including tumour suppressor genes deleted in neuroblastoma and glioma (for example *CAMTA1*), dopamine receptors (*DRD3*) and neurotransmitter transporters (*SLC6A5*, *SLC6A6* and *SLC6A9*). Globally, a gene ontology analysis revealed enrichment for terms relevant to neurogenesis and synaptic function (Supplementary Table 9).

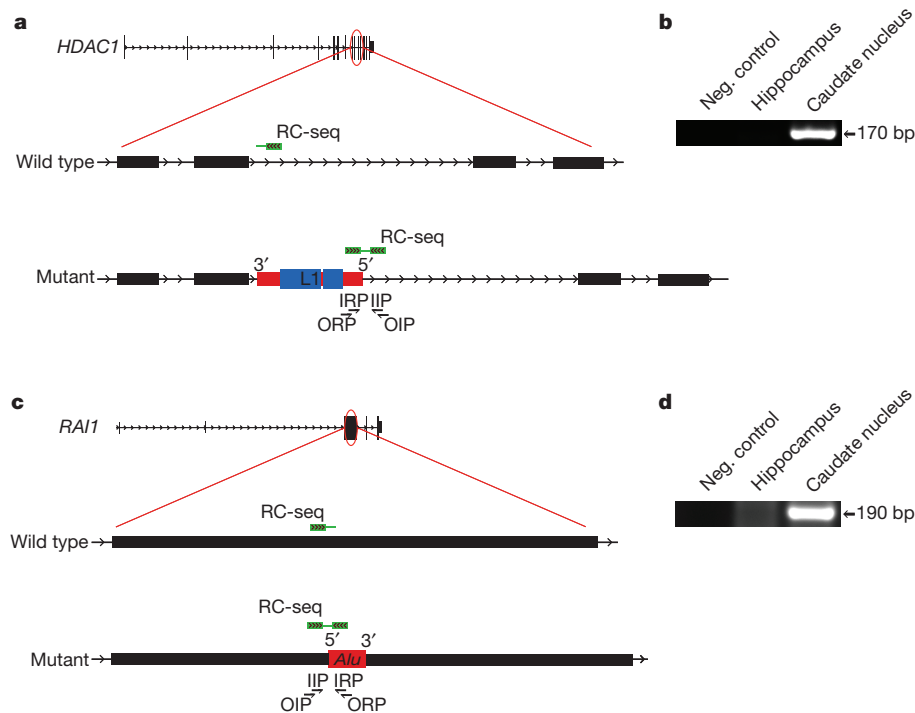
Unlike that for L1, *Alu* retrotransposition has not previously been reported in normal brain cells. However, the L1 transposition machinery is known to *trans*-mobilize *Alu* (ref. 19) and 83.0% of the somatic *Alu* insertions corresponded to the AluY subfamily most active in the human germ line (Supplementary Fig. 3b), making the coincidence of somatic L1 and *Alu* mobilization plausible. On a per-element basis, the observed *Alu* activity was approximately 20-fold lower than that of L1 (Supplementary Results). Thus, it is unlikely that *Alu* CNV would be statistically significant if assayed by TaqMan quantitative PCR<sup>10</sup>. The genomic patterns of *Alu* and L1 insertions were also

different; somatic *Alu* insertions were not overrepresented in introns but were even more common than L1 in exons (Supplementary Table 8). *Alu* exonization is a noted cause of genetic disease<sup>2</sup>. Overall, L1, *Alu* and, to a more limited extent, SVA mobilization produced a large number of insertions that affected protein-coding genes.

Our results indicate that somatic L1 and *Alu* mobilization fundamentally alters the genetic landscape of the human brain, and that retrotransposition is the primary mechanism underlying this phenomenon. By contrast with germline activity<sup>6,16</sup>, somatic insertions disproportionately impacted protein-coding loci. Germline insertions are rarely found in regions where they generate a deleterious phenotype because such mutations are strongly selected against during evolution. Somatic events, on the other hand, are present for one generation and may affect protein-coding loci in a specific environmental context, perhaps being drawn to open chromatin in transcribed regions<sup>15</sup>. Apart from the obvious effects of exonic insertions, intronic events could act as subtle transcriptional 'rheostats'<sup>20</sup> or as *cis*-regulatory elements<sup>21</sup> akin to the IAP insertion responsible for the viable yellow allele of the agouti gene in the mouse<sup>22</sup>.

Several recent studies have catalogued retrotransposon insertions in the human germ line and tumours<sup>6,16,23,24</sup>. Through RC-seq, we have extended these data to the brain and linked somatic retrotransposition to neurobiological genes. For instance, *HDAC1* is a genome-wide transcriptional regulator that controls the canonical L1 promoter<sup>4,25</sup> and is implicated in psychiatric disease and tumorigenesis<sup>26</sup>. Another example highlighted here, *RAI1*, is a transcription factor highly expressed in the brain and previously linked with schizophrenia and Smith–Magenis syndrome<sup>27</sup>. An exonic *Alu* insertion in *RAI1* (Fig. 4c), could therefore have phenotypic consequences.

The hippocampus seems to be predisposed to somatic L1 retrotransposition<sup>10</sup>, which is intriguing given that its subgranular zone is a



**Figure 4 | Discovery of somatic insertions in *HDAC1* and *RAI1*.**

**a**, Alignment of an RC-seq read (green) from donor C caudate nucleus indicated an antisense L1 insertion in intron 9 of *HDAC1*. Nested PCR primers were designed to span the L1 5' terminus, with an initial reaction combining outside retrotransposon primers (ORPs) and insertion site primers (OIPs) and a second reaction using inside retrotransposon primers (IRP) and insertion site primers (IIP). **b**, Amplification of the nested PCR target, confirmed for specificity by capillary sequencing, was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the L1 insertion mobilized from

chromosome 9 and was accompanied by 5' transduction. bp, base pairs.

**c**, Alignment of an RC-seq read pair from donor A caudate nucleus indicated a sense *Alu* insertion in exon 3, and the coding sequence, of *RAI1*. **d**, As for **b**, amplification of the nested PCR target was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the *Alu* insertion mobilized from chromosome 4. We note that L1 and *Alu* elements in **a** and **c** are not drawn to scale, that the L1 open reading frames in **a** are coloured in blue and that the untranslated regions of L1 and *Alu* are coloured in red.

main source of adult neurogenesis<sup>13</sup>. This is also consistent with the hypothesis that L1 retrotransposition is related to neural plasticity<sup>14</sup>. Even more intriguing is the possibility that the APOBEC proteins, which are RNA/DNA-editing enzymes that have expanded under strong positive selection in the primate lineage and been shown to control L1 mobility, may modulate somatic retrotransposition in the brain<sup>28</sup>.

Mutagenesis due to somatic retrotransposition has obvious tumorigenic potential<sup>29</sup> and may have a role in other diseases and biological processes. For example, deletion of the chromatin-remodelling HDAC1 cofactor MECP2<sup>12,25</sup> leads to increased L1 copy number and may inhibit neuronal maturation in Rett syndrome<sup>30</sup>. Somatic mosaicism could also be a factor in neurological dimorphisms seen among discordant monozygotic twins<sup>14</sup>. Future studies may determine whether the overall frequency of somatic retrotransposition varies considerably between individuals, as suggested by our data and previous experiments<sup>10</sup>, and between populations. Ultimately, direct identification of transcripts disrupted by somatic retrotransposition, together with its epigenetic regulation, may provide insights into the molecular processes underlying human cognition, neurodevelopmental disorders and neoplastic transformation.

## METHODS SUMMARY

**Human DNA samples.** Tissues were provided by the Netherlands Brain Bank, Amsterdam. They came from three post-mortem donors with no evidence of neurodegeneration. Pooled human genomic DNA was purchased from Promega. **TaqMan quantitative PCR.** Quantitative PCR experiments were performed with minor modifications to an earlier approach<sup>10</sup>. Quantification included five technical replicates. For each assay, the ratio of L1 ORF2 to  $\alpha$ -satellite repeats was normalized to the ratio obtained for caudate nucleus. Ratios were compared across brain regions by repeated-measures one-way analysis of variance with Bonferroni correction.

**Retrotransposon capture array design.** A NimbleGen Sequence Capture 2.1M Array was customized to contain oligonucleotide probes tiled across the termini of full-length L1, *Alu* and SVA retrotransposons, as well as LTRs intended to act as negative controls. Probes were not filtered for repetitiveness. Typically, eight probes were generated per L1 and SVA retrotransposon and per LTR, and four probes per *Alu* retrotransposon, with a total of 4,885 probes across 875 targeted elements.

**Capture library preparation and sequencing.** DNA sequencing libraries were constructed using an Illumina paired-end kit with substantial modifications (Supplementary Methods). Genomic DNA (2.5  $\mu$ g) was used for each RC-seq library. Amplification based on ligation-mediated PCR was performed before and after hybridization. The average insert size was  $\sim$ 250 nucleotides. Enrichment was confirmed by quantitative PCR against *Alu*. Sequencing was performed by ARK-Genomics, The Roslin Institute, using an Illumina GAIIx instrument.

**Computational analyses.** Paired-end RC-seq reads were mapped to human genome assembly hg19 using SOAP2. Reads where both ends could be aligned to the genome, but not at the same locus, indicated novel retrotransposon insertions. These alignments were corroborated by BLAT, stringently filtered and clustered. Clusters were annotated using published retrotransposon databases<sup>6,16,17</sup> and the NCBI RefSeq database.

Received 14 March; accepted 5 September 2011.

Published online 30 October 2011.

1. Kazanian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
2. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
3. Xing, J. *et al.* Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* **19**, 1516–1526 (2009).
4. Garcia-Perez, J. L. *et al.* Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466**, 769–773 (2010).
5. Yang, N. & Kazanian, H. H. Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13**, 763–771 (2006).
6. Iskow, R. C. *et al.* Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141**, 1253–1261 (2010).

7. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).
8. Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. & Deininger, P. Somatic expression of LINE-1 elements in human tissues. *Nucl. Acids Res.* **38**, 3909–3922 (2010).
9. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41**, 563–571 (2009).
10. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
11. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).
12. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
13. Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nature Med.* **4**, 1313–1317 (1998).
14. Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G. & Gage, F. H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **33**, 345–354 (2010).
15. Cost, G. J., Golding, A., Schlissel, M. S. & Boeke, J. D. Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* **29**, 573–577 (2001).
16. Ewing, A. D. & Kazanian, H. H. Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* **20**, 1262–1270 (2010).
17. Wang, J. *et al.* dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* **27**, 323–329 (2006).
18. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31**, 159–165 (2002).
19. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked *Alu* sequences. *Nature Genet.* **35**, 41–48 (2003).
20. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).
21. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).
22. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
23. Beck, C. R. *et al.* LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170 (2010).
24. Huang, C. R. *et al.* Mobile interspersed repeats are major structural variants in the human genome. *Cell* **141**, 1171–1182 (2010).
25. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
26. Kazantsev, A. G. & Thompson, L. M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nature Rev. Drug Discov.* **7**, 854–868 (2008).
27. Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. & Elsea, S. H. Mutations in *RAI1* associated with Smith-Magenis syndrome. *Nature Genet.* **33**, 466–468 (2003).
28. Mattick, J. S. RNA as the substrate for epigenome-environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition. *Bioessays* **32**, 548–552 (2010).
29. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).
30. Chahrouh, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56**, 422–437 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** J.K.B. is supported by a Wellcome Trust Clinical Fellowship (090385/Z/09/Z) through the Edinburgh Clinical Academic Track. G.J.F. is funded by an Institute Strategic Programme Grant and a New Investigator Award from the British BBSRC (BB/H005935/1) and a C. J. Martin Overseas Based Biomedical Fellowship from the Australian NHMRC (575585). Human brain tissues were provided by the Netherlands Brain Bank to P.H. with ethical consent for them to be used as described in the study.

**Author Contributions** J.K.B., M.W.B., K.R.U., D.J.G., P.R., S.S., P.C. and G.J.F. designed and performed the experiments. J.K.B., T.A.R., F.D.S. and M.F. conducted the computational analyses. P.B., R.T.T., T.C.F., D.A.H., P.H., P.C., J.A.J. and G.J.F. provided resources. S.G. and J.S.M. contributed to the discussion. J.A.J. and G.J.F. invented RC-seq. G.J.F. directed the study, led the bioinformatic analysis and wrote the manuscript. All authors commented on or contributed to the final manuscript.

**Author Information** RC-seq FASTA sequences for brain and blood have been deposited in the NCBI Sequence Read Archive under accession number SRA024401. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.J.F. ([faulknerj@gmail.com](mailto:faulknerj@gmail.com)).



# Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination

Kerstin Berer<sup>1</sup>, Marsilius Mues<sup>1</sup>, Michail Koutrolos<sup>1</sup>, Zakeya Al Rasbi<sup>1</sup>, Marina Boziki<sup>1</sup>, Caroline Johner<sup>2</sup>, Hartmut Wekerle<sup>1</sup> & Gurumoorthy Krishnamoorthy<sup>1</sup>

Active multiple sclerosis lesions show inflammatory changes suggestive of a combined attack by autoreactive T and B lymphocytes against brain white matter<sup>1</sup>. These pathogenic immune cells derive from progenitors that are normal, innocuous components of the healthy immune repertoire but become autoaggressive upon pathological activation. The stimuli triggering this autoimmune conversion have been commonly attributed to environmental factors, in particular microbial infection<sup>2</sup>. However, using the relapsing–remitting mouse model of spontaneously developing experimental autoimmune encephalomyelitis<sup>3</sup>, here we show that the commensal gut flora—in the absence of pathogenic agents—is essential in triggering immune processes, leading to a relapsing–remitting autoimmune disease driven by myelin-specific CD4<sup>+</sup> T cells. We show further that recruitment and activation of autoantibody-producing B cells from the endogenous immune repertoire depends on availability of the target autoantigen, myelin oligodendrocyte glycoprotein (MOG), and commensal microbiota. Our observations identify a sequence of events triggering organ-specific autoimmune disease and these processes may offer novel therapeutic targets.

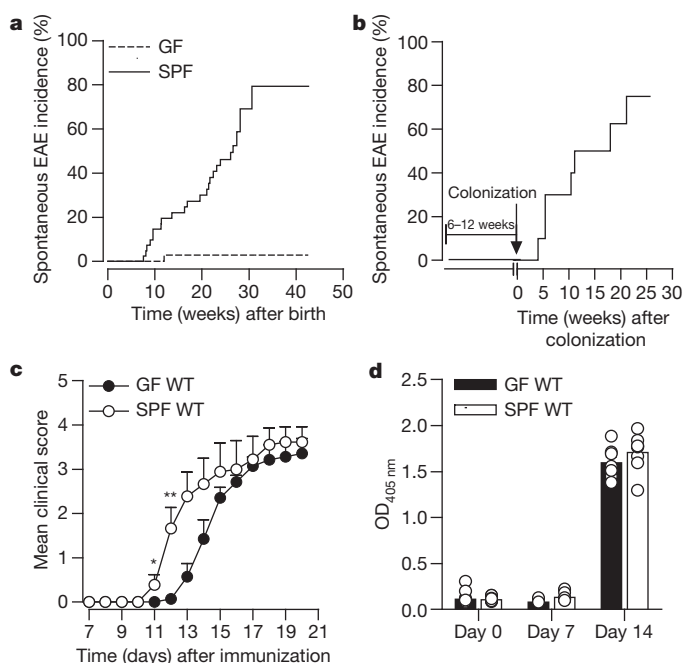
The relapsing–remitting (RR) mouse model uses transgenic SJL/J mice expressing, in a large proportion of their CD4<sup>+</sup> T cells, a transgenic T-cell antigen receptor (TCR) recognizing MOG peptide 92–106 in the context of MHC class II, I-A<sup>s</sup>. These mice spontaneously develop experimental autoimmune encephalomyelitis (EAE) with successive disease bouts that often affect different central nervous system (CNS) tissues. The disease is initiated by the transgenic CD4<sup>+</sup> T cells, which first infiltrate the CNS, and by MOG-autoantibody-producing B cells recruited from the natural immune repertoire<sup>3</sup>.

Whereas in our facility close to 80% of RR mice developed spontaneous EAE within 3–8 months of age, the rate was variable in other institutions, with spontaneous EAE incidences ranging from 35–90% (unpublished data). This recalled previous investigations that also observed that the frequency of spontaneous EAE in myelin-specific TCR transgenic mice varied in different breeding centres<sup>4</sup>. Because our mice were reared under specific pathogen-free (SPF) conditions, we tested the possible contributions of the non-pathogenic commensal flora to the triggering of a spontaneous CNS-specific autoimmune disease.

We first compared the incidence of spontaneous EAE between RR mice housed under SPF and completely germ-free conditions. The differences were marked. Whereas, as reported before, most SPF-bred RR mice came down with EAE within 3–8 months<sup>3</sup>, germ-free RR mice remained fully protected throughout their life (Fig. 1a). As the commensal microbiota have a central function in driving the correct development of the immune system<sup>5</sup>, the absence of spontaneous EAE in germ-free RR mice may have reflected a general immune deficiency due to missing microbial stimuli. However, two observations argue against a profound and irreversible non-reactivity. First, RR mice, which had been germ free (and disease free) for 6–12 weeks, promptly developed EAE when re-colonized with conventional commensal microbiota (Fig. 1b). Mono-colonization with segmented filamentous

bacteria (SFB), which restored autoimmunity in another mouse model, was of low efficiency (unpublished data). This suggests that the immune system of germ-free mice had grown efficient enough to mount a full autoimmune attack within a relatively brief period of time. Second, the basic immune competence of germ-free animals was confirmed by active immunization of germ-free wild-type SJL/J mice with recombinant MOG (rMOG) in complete Freund's adjuvant (CFA). In accord with one previous report<sup>6</sup>, although not with another more recent one<sup>7</sup>, all immunized germ-free mice developed EAE like their SPF counterparts, although with some delay (Fig. 1c), and transfer of pre-activated T cells induced comparable EAE in both germ-free and SPF mice (Supplementary Table 1). Moreover, germ-free and SPF SJL/J mice immunized with rMOG produced comparable levels of anti-MOG antibodies in their serum (Fig. 1d).

Recent studies established that components of the commensal microbiota profoundly shape the gut-associated lymphatic tissue (GALT),



**Figure 1 | Commensal microbiota are required for the development of spontaneous EAE.** **a**, Incidence of spontaneous EAE in a cohort of RR mice housed in germ-free (GF;  $n = 35$ ) or SPF ( $n = 41$ ) conditions. **b**, Incidence of spontaneous EAE in germ-free RR mice ( $n = 10$ ) re-colonized with conventional flora from SPF mice. **c**, Delayed EAE onset in germ-free wild-type (GF WT) SJL/J mice immunized with rMOG/CFA. Mean EAE scores ( $\pm$  s.e.m.) of germ-free ( $n = 7$ ) and SPF ( $n = 8$ ) SJL/J mice are shown. \* $P < 0.05$ ; \*\* $P < 0.01$  (two-way ANOVA). **d**, Germ-free and SPF wild-type SJL/J mice produce similar levels of anti-MOG antibodies after immunization. Each circle represents an individual mouse and bars depict mean  $\pm$  s.e.m. Panels c and d represent two individual experiments.

<sup>1</sup>Department of Neuroimmunology, Max Planck Institute of Neurobiology, 82152 Martinsried, Germany. <sup>2</sup>Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany.

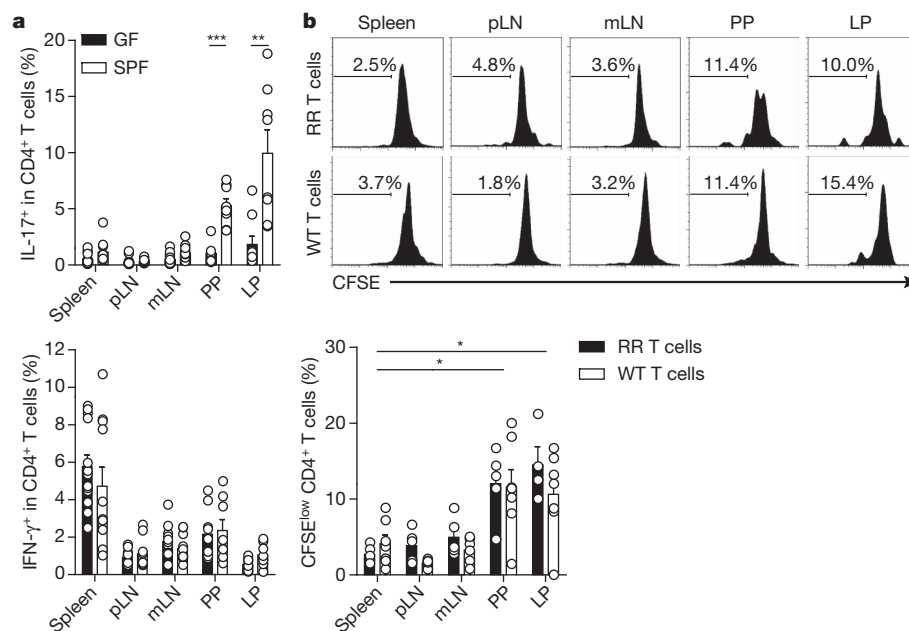
some supporting differentiation of interleukin (IL)-17-producing  $T_H17$  cells<sup>8–10</sup> and others the generation of regulatory T lymphocytes ( $T_{reg}$ )<sup>11,12</sup>. We also found a marked deficit of  $T_H17$ -like cells in germ-free mice (Fig. 2a), which was most pronounced in T cells intimately connected to the intestinal wall, lamina propria T cells and in Peyer's patch but not in mesenteric lymph node populations. There were no notable changes in remote organs, such as spleen or pooled inguinal and axillary lymph nodes (Fig. 2a). Frequencies of IFN- $\gamma$ -, TNF- $\alpha$ - and IL-10-producing  $CD4^+$  T cells were comparable between germ-free and SPF RR mice (Fig. 2a and Supplementary Fig. 1). Apart from a minor increase in the frequency of  $CD4^+$  T cells in the spleen of germ-free RR mice and a reduction of the T cells expressing lower levels of T-cell receptor ( $TCR\alpha\beta^{low}$ ) (activated T cells) in the lamina propria (Supplementary Figs 2 and 3), the proportions of most lymphoid cell types examined, including  $Foxp3^+$   $T_{reg}$  cells,  $CD8^+$  T cells,  $TCR\gamma\delta^+$  cells, B cells,  $CD11b^+$  macrophages,  $CD11c^+$  dendritic cells, natural killer (NK) cells and  $Gr1^+$  granulocytes were unchanged (Supplementary Fig. 2). Of note, although in the spleen the microbial colonization status did not affect cellular composition, it definitely impinged on cytokine production of splenic immune cells. As in MOG-immunized germ-free C57BL/6 mice<sup>7</sup>, germ-free RR mouse spleen cells released lower levels of IL-17 than their SPF counterparts upon MOG antigen or anti-CD3 monoclonal antibody stimulation, and in addition they showed reduced secretion of IFN- $\gamma$ . Re-colonization of germ-free mice not only restored T-cell cytokine production capacity but even led to overshooting reactions (Supplementary Fig. 4).

The commensal microbiota could act on MOG-specific T cells either via microbial structures mimicking MOG epitopes<sup>13</sup> or through innate immune signals creating a particular inflammatory milieu<sup>11,14</sup>. In an attempt to probe a potential MOG-specific mimicry response, we transferred carboxyfluorescein succinimidyl ester (CFSE)-labelled TCR transgenic or wild-type T cells into SPF wild-type mice and tested their proliferative responses in the gut. Proliferation rates of transgenic and polyspecific wild-type T cells in the GALT were equally high, whereas in the remote spleen of the same recipients the responses remained hardly detectable (Fig. 2b). Further, the microbial signals

seem to act persistently on local T cells. Transient depletion of gut flora by short-term antibiotic treatment significantly reduced the proliferation of T cells in the lamina propria, but not in spleen, pooled lymph nodes, mesenteric lymph nodes and Peyer's patches (Supplementary Fig. 5).

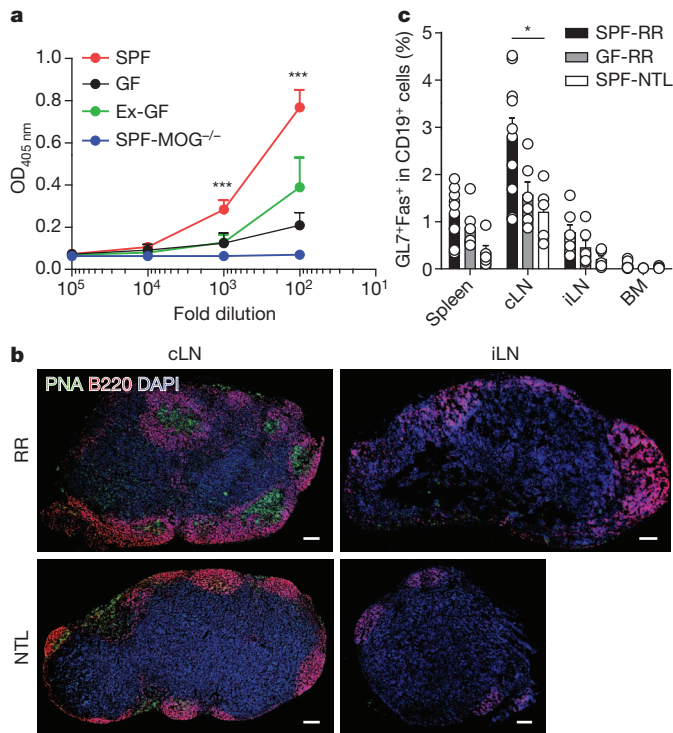
Activation of MOG-specific T cells in the GALT is necessary for the development of EAE in RR mice, but not sufficient. Full clinical EAE requires the participation of MOG-reactive B lymphocytes. We proposed that in RR mice, transgenic pathogenic T cells select the autoimmune B cells from the native B-cell repertoire and drive them to proliferate and release autoantibodies of IgG classes<sup>3</sup>. Indeed, we now found that germ-free RR mice, which, owing to missing microbial stimuli, lack activated autoimmune T cells, produced only low doses of anti-MOG autoantibodies. The autoantibody production was promptly increased in germ-free mice upon re-colonization (Fig. 3a). This response could involve antigenic mimicry at the B-cell level between MOG and epitopes on commensal microbes, reminiscent of Sydenham's chorea—the CNS manifestation of rheumatic fever—in which streptococcal antigens mimic neuronal B-cell epitopes<sup>15</sup>. However, this was not the case in spontaneous RR mouse EAE. We discovered that production of demyelinating autoantibodies critically depended on the expression of the target myelin autoantigen, MOG. RR mice deficient in MOG ( $RR \times MOG^{-/-}$ ), due to a transgenic mutation of the *Mog* gene<sup>16</sup>, failed to develop anti-MOG autoantibody titres despite their normal microbial status (Fig. 3a). Importantly, our data show that exogenous MOG injected into SPF  $RR \times MOG^{-/-}$  mice via MOG in CFA readily induced anti-MOG antibodies (Supplementary Fig. 6).

Recruitment and activation of antigen-specific B cells involves signals by local helper T cells and surrounding stroma cells, which together drive the resting B cell into a germinal centre, where it undergoes proliferation, immunoglobulin class switching and somatic hypermutation<sup>17</sup>. Binding of the cognate antigen to the B-cell receptor has a central role in these processes. MOG-specific B cells could be recruited either in the CNS tissue via locally produced MOG material<sup>18</sup>, or in CNS draining peripheral lymph nodes (deep cervical lymph



**Figure 2 | Effect of microbiota on T-cell activation and their cytokine profiles in the GALT.** **a**, Impaired  $T_H17$  differentiation in germ-free (GF) RR mice. Frequencies of IL-17- or IFN- $\gamma$ -producing T cells from the indicated organs of GF and SPF RR mice are shown. LP, lamina propria; mLN, mesenteric lymph nodes; pLN, pooled inguinal and axillary lymph nodes; PP, Peyer's patches.  $n = 8$ –13 mice per group. Data were pooled from four independent

experiments. \*\*\* $P = 0.0002$ ; \*\* $P = 0.0025$  (Mann–Whitney U test). **b**, Activation of T cells by commensal flora. Shown are the frequencies of CFSE<sup>low</sup>  $CD4^+$  cells in the indicated organs of mice that received CFSE-labelled  $CD4^+$  T cells. Each circle represents an individual mouse and bars depict mean  $\pm$  s.e.m.  $n = 4$ –7 mice per group. Data represent two individual experiments. \* $P < 0.05$  (Mann–Whitney U test).

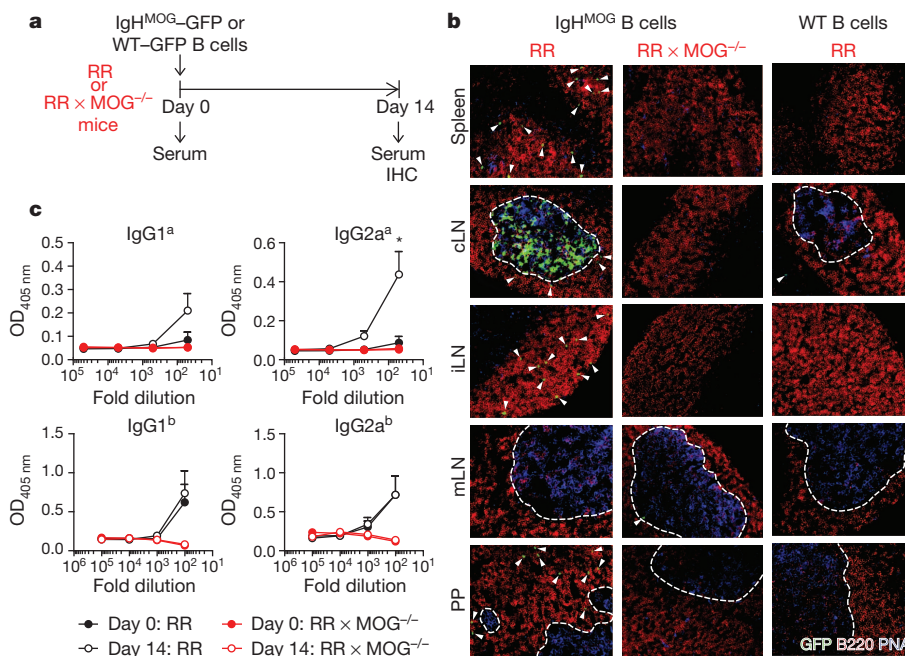


**Figure 3 | B-cell recruitment is impaired in germ-free RR mice.** **a**, Reduced production of MOG-specific IgG2a antibodies in sera of germ-free (GF) RR mice. SPF ( $n = 15$ ); germ-free ( $n = 24$ ); germ-free mice re-colonized with conventional commensal microbiota (Ex-GF;  $n = 7$ ); or SPF-MOG<sup>-/-</sup> ( $n = 13$ ) RR mice. Error bars indicate s.e.m. \*\*\* $P < 0.001$  (Kruskal–Wallis test). **b**, **c**, Spontaneously formed germinal centre B cells are enriched in the cervical lymph nodes (cLN) of RR mice. **b**, Immunofluorescence staining of cervical lymph nodes or inguinal lymph node (iLN) sections of RR and NTL mice. Peanut agglutinin (PNA) (green; germinal centre), B220 (red; B cells) and 4',6-diamidino-2-phenylindole (DAPI) (blue; cell nuclei). Representative data of 4–5 individual mice are shown. Scale bars, 100  $\mu$ m. **c**, Flow cytometric analysis of GL7<sup>+</sup> Fas<sup>+</sup> germinal centre B cells from spleen, cervical lymph nodes, inguinal lymph nodes and bone marrow (BM). Each circle represents an individual mouse and bars represent mean  $\pm$  s.e.m.  $n = 5$ –10 per group. Data were pooled from three independent experiments. \* $P < 0.05$  (Mann–Whitney U test).

nodes) with MOG imported from the CNS via lymphatic vessels<sup>19</sup>. Our observations favour the latter alternative. Prior to the onset of clinical symptoms in SPF RR mice we found some scattered B cells in CNS infiltrates (Supplementary Fig. 7) but no follicle-like aggregates or follicular markers (data not shown). However, there were conspicuous changes in the cervical lymph nodes. Starting from the age of 3 weeks, cervical lymph nodes contained clearly delineated germinal centres (Fig. 3b). Furthermore, germinal centres and increased frequencies

of GL7<sup>+</sup> and Fas<sup>+</sup> germinal centre B cells were restricted to cervical lymph nodes of RR mice, but were significantly reduced in age-matched germ-free and non-transgenic littermates (NTL) (Fig. 3b, c).

Germinal centres are attractive milieus for B cells, provided they contain appropriate antigenic material and competent T-helper cells<sup>20</sup>. To investigate whether RR mouse cervical lymph nodes offer both prerequisites to MOG-reactive B cells, we transferred GFP-labelled, MOG-reactive B cells expressing the H chain of a MOG monoclonal antibody (IgH<sup>MOG</sup>)<sup>21</sup> into hosts with a distinct allotype and traced their homing behaviour (Fig. 4a). When examined 14 days after transfer, MOG-specific B cells were found densely packed within germinal centres of cervical lymph nodes (Fig. 4b). Further, donor IgH<sup>MOG</sup> B cells switched anti-MOG antibodies to IgG2a isotypes (Fig. 4c). However, GFP-labelled polyclonal wild-type B cells failed to accumulate in any of the lymph nodes of RR mice (Fig. 4b), and IgH<sup>MOG</sup> B cells neither homed to the cervical lymph nodes of wild-type or RR  $\times$  MOG<sup>-/-</sup> mice (Fig. 4b and Supplementary Fig. 8) nor produced class-switched anti-MOG antibodies (Fig. 4c and Supplementary Fig. 8c). Collectively, these data indicate an ongoing MOG-specific germinal centre reaction, which is critically dependent on the expression of MOG, in the cervical lymph nodes of RR mice.



**Figure 4 | MOG-specific B cells home to the germinal centre of brain draining cervical lymph nodes.** **a**, Schematic representation of experimental set up. IHC, immunohistochemistry. **b**, MOG-specific B cells home to the germinal centre of brain draining cervical lymph nodes. PNA (blue; germinal centre) and B220 (red; B cells). Arrowheads indicate transferred GFP<sup>+</sup> B cells. Dotted lines define boundaries of germinal centre. Representative staining of

two independent experiments are shown.  $n = 4$ –5 mice per group. Magnification:  $\times 20$ . **c**, Transgenic MOG-specific B cells spontaneously switch isotype in RR but not in RR  $\times$  MOG<sup>-/-</sup> mice. Titres of donor (a allotype) and recipient (b allotype) anti-MOG antibodies were measured. Error bars represent s.e.m.  $n = 4$ –5 mice per group. Data were pooled from three independent experiments. \* $P < 0.05$  (Mann–Whitney U test).



This is the first report, to our knowledge, describing the sequential roles of the intact commensal gut flora and of myelin autoantigen in the initiation of a complex spontaneous demyelinating autoimmune disease. We propose a two-phase scenario that starts out in the GALT with expanding and activating CNS autoreactive T cells, which then recruit autoantibody-producing B cells. Together the autoimmune T and B cells trigger a demyelinating encephalomyelitis, which in the RR SJL/J mouse takes a relapsing–remitting course, very similar to early human multiple sclerosis.

Our findings are of direct relevance to multiple sclerosis, the pathogenesis of which is presently hotly debated. Some propose primary changes in the CNS target as the initiating process<sup>22</sup>, whereas others suggest that pathogenesis originates in the immune system<sup>23</sup>. Our present data support the latter concept. It is tempting to extend our finding of the gut origin of experimental CNS autoimmunity to human multiple sclerosis. There is now emerging evidence implicating gut microbiota in the starting phase of human autoimmune diseases. Besides inflammatory bowel diseases, in which bacteria may act on local tissue directly as well as indirectly<sup>24</sup>, inflammatory diseases with remote tissues affected seem to be modulated by the gut environment; for example, in rheumatoid arthritis<sup>25</sup> and type 1 diabetes mellitus<sup>26</sup>. In multiple sclerosis, evidence for commensal microbial contributions has remained less clear, so far. Dietary risk factors have been suggested to have a role<sup>27</sup>, and may contribute to a conspicuous increase of multiple sclerosis prevalence in Asian countries, like Japan, which has been ascribed to the spreading of a ‘westernized’ lifestyle<sup>28</sup>. It will be of interest now to search for the composition of intestinal microbiota associated with an increased susceptibility to multiple sclerosis, and this may provide a conceptual basis for exploring new, non-invasive treatment strategies.

## METHODS SUMMARY

**Mice.** Germ-free animals were re-derived from SJL/J anti-MOG TCR transgenic RR mice and kept germ free at the animal facility of the Max Planck Institute of Immunobiology and Epigenetics. Mice were re-colonized by housing in bedding material pre-conditioned by SPF mice.

**Cell purification, flow cytometry and adoptive transfers.** Single-cell suspensions were prepared from spleen, lymph nodes, Peyer’s patches and lamina propria by enzymatic digestion or mechanical disruption. Untouched T cells and B cells were purified using negative isolation kits (R&D Systems). Cells were stained with fluorochrome-labelled antibodies and acquired on FACSCalibur (BD Biosciences). Data were analysed using FlowJo (TreeStar) software. CFSE-labelled T cells or GFP<sup>+</sup> B cells were injected intravenously into SPF mice. CFSE<sup>low</sup> T cells were quantified by FACS 3 days after transfer. Localization of GFP<sup>+</sup> B cells was documented by immunofluorescence after 2 weeks.

**Immunofluorescence.** Sections of immune organs were stained with PNA (Vector Laboratories), anti-mouse B220 (BD Biosciences) and DAPI (Invitrogen). Images were obtained with a fluorescence microscope (Axiovert 200M; Carl Zeiss) and processed with MetaMorph 7.7 Software and Adobe Photoshop CS4.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 June; accepted 12 September 2011.**

**Published online 26 October 2011.**

1. Franciotta, D. *et al.* B cells and multiple sclerosis. *Lancet Neurol.* **79**, 852–858 (2008).
2. Ascherio, A. & Munger, K. L. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann. Neurol.* **61**, 288–299 (2007).
3. Pöhlner, B. *et al.* Spontaneous relapsing–remitting EAE in the SJL/J mouse: MOG-reactive transgenic T cells recruit endogenous MOG-specific B cells. *J. Exp. Med.* **206**, 1303–1316 (2009).
4. Goverman, J. *et al.* Transgenic mice that express a myelin basic protein-specific T cell receptor develop spontaneous autoimmunity. *Cell* **72**, 551–560 (1993).
5. Hill, D. A. & Artis, D. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu. Rev. Immunol.* **28**, 623–667 (2010).

6. Lampropoulou, V. *et al.* TLR-activated B cells suppress T cell-mediated autoimmunity. *J. Immunol.* **180**, 4763–4773 (2008).
7. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108**, 4615–4622 (2011).
8. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
9. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
10. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
11. Hall, J. A. *et al.* Commensal DNA limits regulatory T cell conversion and is a natural adjuvant of intestinal immune responses. *Immunity* **29**, 637–649 (2008).
12. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
13. Ercolini, A. M. & Miller, S. D. Molecular mimics can induce novel self peptide-reactive CD4<sup>+</sup> T cell clonotypes in autoimmune disease. *J. Immunol.* **179**, 6604–6612 (2007).
14. Feng, T. *et al.* Microbiota innate stimulation is a prerequisite for T cell spontaneous proliferation and induction of experimental colitis. *J. Exp. Med.* **207**, 1321–1332 (2010).
15. Kirvan, C. A., Cox, C. J., Swedo, S. E. & Cunningham, M. W. Tubulin is a neuronal target of autoantibodies in Sydenham’s chorea. *J. Immunol.* **178**, 7412–7421 (2007).
16. Delarasse, C. *et al.* Myelin/oligodendrocyte glycoprotein-deficient (MOG-deficient) mice reveal lack of immune tolerance to MOG in wild-type mice. *J. Clin. Invest.* **112**, 544–553 (2003).
17. Cyster, J. G. B cell follicles and antigen encounters of the third kind. *Nature Immunol.* **11**, 989–996 (2010).
18. McMahon, E. J. *et al.* Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. *Nature Med.* **11**, 335–339 (2005).
19. Cserr, H. F., Harling-Berg, C. J. & Knopf, P. M. Drainage of brain extracellular fluid into blood and deep cervical lymph and its immunological significance. *Brain Pathol.* **2**, 269–276 (1992).
20. Schwickert, T. A., Alabyev, B., Manser, T. & Nussenzweig, M. C. Germinal center reutilization by newly activated B cells. *J. Exp. Med.* **206**, 2907–2914 (2009).
21. Litzengruber, T. *et al.* B lymphocytes producing demyelinating autoantibodies: development and function in gene-targeted transgenic mice. *J. Exp. Med.* **188**, 169–180 (1998).
22. Barnett, M. H., Henderson, A. P. D. & Prineas, J. W. The macrophage in MS: just a scavenger after all? Pathology and pathogenesis of the acute MS lesion. *Ms* **12**, 121–132 (2006).
23. Hohlfeld, R. Multiple sclerosis: human model for EAE? *Eur. J. Immunol.* **39**, 2036–2039 (2009).
24. Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
25. Vahtovuo, J. *et al.* Fecal microbiota in early rheumatoid arthritis. *J. Rheumatol.* **35**, 1500–1505 (2008).
26. Knip, M. *et al.* Dietary intervention in infancy and later signs of  $\beta$ -cell autoimmunity. *N. Engl. J. Med.* **363**, 1900–1908 (2010).
27. Lauer, K. Diet and multiple sclerosis. *Neurol. (Tokyo)* **49** (suppl. 2), S55–S61 (1997).
28. Kira, J.-I. *et al.* Changes in the clinical phenotypes of multiple sclerosis during the past 50 years in Japan. *J. Neurol. Sci.* **166**, 53–57 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank I. Arnold-Ammer, N. Reißer and L. Penner for technical assistance, and M. Pfunder, U. Stauffer, C. Hornung and N. Joswig for maintaining our germ-free colony and technical support. We are much obliged to R. Kemler for his support. This work was funded by SFB 571 (Project B6), the German Competence Network on Multiple Sclerosis (KKNMS), ARSEP (France), and by the Max Planck Society. K.B. is supported by a fellowship from ARSEP. Z.A.R. is supported by a PhD fellowship from the Emirates Foundation. M.B. receives a fellowship from the Hellenic Neurological Society.

**Author Contributions** K.B., H.W. and G.K. designed experiments and wrote the manuscript with input from co-authors. K.B. performed most of the experiments. M.M., M.K. and M.B. performed flow cytometry experiments or assisted in experiments. Z.A.R. performed flow cytometry and immunofluorescence staining for brain infiltrates. C.J. supervised the maintenance of germ-free mouse colony and colonization experiments.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.W. ([hwekerle@neuro.mpg.de](mailto:hwekerle@neuro.mpg.de)) or G.K. ([guru@neuro.mpg.de](mailto:guru@neuro.mpg.de)).

## METHODS

**Mice and colonization.** Wild-type SJL/J, RR and RR  $\times$  MOG<sup>-/-</sup> SJL/J and IgH<sup>MOG</sup>  $\times$  actin-GFP and actin-GFP SJL/J mice were bred at the animal facility of the Max Planck Institute of Neurobiology. Germ-free RR mice were obtained by transferring embryos, isolated by sterile hysterectomy on embryonic day (E) 18.5, to sterile breeding conditions and by fostering on germ-free foster mothers. They were bred and maintained in positive-pressure plastic isolators and provided with  $\gamma$ -irradiated commercial rodent diet and autoclaved water at the animal facility of the Max Planck Institute of Immunobiology and Epigenetics. Fecal samples were routinely cultured in standard I-Bouillon and examined for contamination. In addition, mice from the colony were screened bi-annually according to FELASA health monitoring recommendations. For re-colonization experiments, germ-free mice were placed in cages with bedding material pre-conditioned by conventional (SPF) mice. All animal procedures were in accordance with the guidelines of the Committee on Animals of the Max Planck Institute of Neurobiology and with a license from the Regierung von Oberbayern.

**Active induction of EAE.** Mice were immunized subcutaneously with 200  $\mu$ g rMOG emulsified in Freund's adjuvant supplemented with 5 mg ml<sup>-1</sup> *Mycobacterium tuberculosis* (strain H37Ra; Difco). On days 0 and 2 after immunization, 200 ng of pertussis toxin (List Biological Laboratories) were injected intraperitoneally. Clinical scoring of EAE was done as published<sup>3</sup>.

**Antibiotic treatment.** For short-term antibiotic treatment, 8-week-old wild-type SJL/J mice were treated for 7 days with 1 g l<sup>-1</sup> of metronidazole (Sigma), 1 g l<sup>-1</sup> of neomycin (Sigma) and 0.5 g l<sup>-1</sup> of vancomycin (AppliChem) in their drinking water.

**CFSE labelling and adoptive transfer.** Splenocytes from RR or wild-type mice were labelled at 37 °C for 10 min with 5  $\mu$ M CFSE (Invitrogen) in PBS containing 1% fetal bovine serum (FBS). Cells were washed twice in ice-cold PBS and subsequently CD4<sup>+</sup> T cells were isolated using a mouse CD4<sup>+</sup> T-cell isolation kit (R&D Systems).  $5 \times 10^6$  CFSE-labelled CD4<sup>+</sup> T cells were injected intravenously into wild-type SJL/J mice.

**B-cell isolation and adoptive transfer.** B cells were isolated from spleens using a mouse B-cell isolation kit (R&D Systems). B cells were enriched to >90% purity, as confirmed by flow cytometry.  $10 \times 10^6$  purified IgH<sup>MOG</sup>-GFP or wild-type-GFP B cells were intravenously injected into RR, wild-type or RR  $\times$  MOG<sup>-/-</sup> mice.

**Proliferation assay.** For the proliferation assay,  $2 \times 10^5$  splenocytes were cultured in the presence of various concentrations of rMOG or anti-CD3 antibody (BD Pharmingen), as indicated. Proliferative response was measured by the incorporation of [<sup>3</sup>H]-thymidine (1  $\mu$ Ci well<sup>-1</sup>) during the last 16 h of a 72 h culture period. Proliferation assays were performed in triplicates.

**Cell isolation and flow cytometry.** Single-cell suspensions were prepared from spleen, pooled peripheral lymph nodes (axillary plus inguinal), or individual lymph nodes (cervical and inguinal), mesenteric lymph nodes or Peyer's patches by mechanical disruption via forcing through 40- $\mu$ m cell strainers (BD Biosciences). For the isolation of lamina propria lymphocytes, small intestine was collected in ice-cold HBSS buffered with 15 mM HEPES. After careful removal

of Peyer's patches, fatty tissue and fecal contents, the intestine was opened longitudinally and cut into small pieces. The intestinal fragments were washed three times for 15 min with stirring (300 r.p.m.) in HBSS containing 5 mM EDTA, 15 mM HEPES and 10% FBS. Next, intestinal pieces were washed once for 5 min with stirring in RPMI containing 15 mM HEPES and 10% FBS, followed by an incubation step at 37 °C with stirring (500 r.p.m.) in RPMI with 15 mM HEPES, 10% FBS and 100 U ml<sup>-1</sup> Collagenase VII (Sigma). The digested tissue was washed once in RPMI with 15 mM HEPES and 10% FBS, before the lamina propria lymphocytes were subjected to FACS analysis. CNS infiltrating cells were purified by Percoll gradient centrifugation as described<sup>3</sup>. For detection of cell surface markers, cells were stained in FACS buffer (PBS containing 1% BSA and 0.1% Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>) with fluorochrome-labelled monoclonal antibodies: PerCP-conjugated anti-CD4 (RM4-5), PerCP-Cy5.5-conjugated anti-B220 (RA3-6B2), PE- and APC-conjugated anti-CD19 (1D3), APC-conjugated anti-CD8 $\alpha$  (53-6.7), PE-conjugated anti-TCR $\gamma\delta$  (eBioGL3), PE-conjugated anti-CD11b (M1/70), APC-conjugated anti-NKp46 (29A1.4), FITC-conjugated anti-CD11c (HL3), biotin-conjugated anti-Gr1 (RB6-8C5), FITC-conjugated anti-CD45.1 (A20), FITC-conjugated anti-V $\beta$ 4 (KT4), PE-conjugated anti-V $\alpha$ 8.3 (B21.14), FITC-conjugated anti-GL7, PE-conjugated anti-Fas (Jo2) and PE-conjugated streptavidin. For intracellular cytokine staining, cells were activated with 50 ng ml<sup>-1</sup> PMA (Sigma) and 500 ng ml<sup>-1</sup> ionomycin (Sigma) in the presence of 5  $\mu$ g ml<sup>-1</sup> brefeldin A (Sigma) for 4 h at 37 °C. After surface staining, cells were fixed and permeabilized in 4% paraformaldehyde/0.1% saponin in HEPES-buffered HBSS and stained intracellularly using the following antibodies: PE-conjugated anti-IL17 (TC11-18H10), APC-conjugated anti-IFN- $\gamma$  (XMG1.2), APC-conjugated anti-TNF- $\alpha$  (MP6-XT22), PE-conjugated anti-IL-10 (JES5-16E3) and APC-conjugated anti-FoxP3 (FJK-16 s). All antibodies were purchased from BD Pharmingen or eBioscience. Cells were acquired on a FACSCalibur (BD Biosciences) and analysis was performed using FlowJo (TreeStar) software.

**ELISA.** Serum titres of anti-MOG antibodies were quantified as previously described<sup>3</sup>. Cytokine levels in cell culture supernatants were determined with antibody pairs for IFN- $\gamma$  (BD Biosciences) or IL-17 (eBioscience).

**Immunofluorescence.** Organs were fixed in PBS with 4% paraformaldehyde and cryoprotected in PBS plus 30% sucrose before embedding in OCT medium (A. Hartenstein). Cryostat sections (10  $\mu$ m in thickness) of spleen, lymph nodes and brains were fixed in acetone. Sections were blocked with PBS and 5% BSA before being stained in a humidified chamber. The following antibodies were used for staining: biotin-conjugated anti-CD4 (BD Pharmingen), purified anti-B220 (BD Pharmingen), biotin-conjugated PNA (Vector Laboratories), Alexa Fluor 568-conjugated anti-rat IgG (Invitrogen), Alexa Fluor 488-conjugated Streptavidin (Invitrogen), APC-conjugated streptavidin (eBioscience) and DAPI (Invitrogen). Images were obtained with a fluorescence microscope (Axiovert 200M; Carl Zeiss) and processed with MetaMorph 7.7 Software and Adobe Photoshop CS4.

**Statistical analysis.** GraphPad Prism 5 (GraphPad Software) was used for all statistical analysis. *P* values <0.05 were considered to be significant.

# Dendritic cells control lymphocyte entry to lymph nodes through high endothelial venules

Christine Moussion<sup>1,2</sup> & Jean-Philippe Girard<sup>1,2</sup>

While patrolling the body in search of foreign antigens, naive lymphocytes continuously circulate from the blood, through the lymph nodes, into the lymphatic vessels and back to the blood<sup>1,2</sup>. This process, called lymphocyte recirculation, provides the body with effective immune surveillance for foreign invaders and for alterations to the body's own cells. However, the mechanisms that regulate lymphocyte recirculation during homeostasis remain incompletely characterized. Here we show that dendritic cells (DCs), which are well known for their role in antigen presentation to T lymphocytes<sup>3</sup>, control the entry of naive lymphocytes to lymph nodes by modulating the phenotype of high endothelial venules (HEVs), which are blood vessels specialized in lymphocyte recruitment<sup>2,4,5</sup>. We found that *in vivo* depletion of CD11c<sup>+</sup> DCs in adult mice over a 1-week period induces a reduction in the size and cellularity of the peripheral and mucosal lymph nodes. In the absence of DCs, the mature adult HEV phenotype reverts to an immature neonatal phenotype, and HEV-mediated lymphocyte recruitment to lymph nodes is inhibited. Co-culture experiments showed that the effect of DCs on HEV endothelial cells is direct and requires lymphotoxin- $\beta$ -receptor-dependent signalling. DCs express lymphotoxin, and DC-derived lymphotoxin is important for lymphocyte homing to lymph nodes *in vivo*. Together, our results reveal a previously unsuspected role for DCs in the regulation of lymphocyte recirculation during immune surveillance.

Lymphocyte entry to lymph nodes is initiated by the lymphocyte homing receptor L-selectin, which mediates lymphocyte rolling along HEV walls<sup>2,4-6</sup>. In contrast to the endothelium that lines other vessels, HEV endothelial cells in lymph nodes have a plump, almost cuboidal morphology<sup>4,5</sup>, and they express high levels of sulphated carbohydrate ligands for L-selectin, which are recognized by the monoclonal antibody MECA-79 (refs 6–9) and are synthesized by the HEV-specific enzymes GlcNAc6ST-2 (a sulphotransferase)<sup>8,9</sup> and FucT-VII (a fucosyltransferase)<sup>10</sup>. Studies carried out in rodents have shown that, when the peripheral lymph nodes (PLNs) are deprived of afferent lymph, the HEVs lose their characteristics and their ability to support lymphocyte traffic<sup>11–14</sup>. Together with data obtained from human studies<sup>15</sup>, these results indicate that the lymphoid tissue microenvironment is crucial for the maintenance of HEV characteristics during homeostasis<sup>4,15</sup>. However, the cell types involved have not been defined.

To address the potential role of DCs in the maintenance of HEVs, we took advantage of the CD11c–diphtheria toxin receptor (*Cd11c-DTR*) transgenic mouse model<sup>16</sup>, which allows *in vivo* depletion of CD11c<sup>+</sup> DCs in adult mice after treatment with diphtheria toxin (DTX). Because downregulation of the HEV phenotype occurs 1 week after occlusion of the afferent lymphatics<sup>12</sup>, we decided to deplete the DCs over 1 week by injecting DTX every 2 days (Fig. 1a and Supplementary Fig. 1). After 8 days of DC depletion, we observed a reduction in the size and cellularity of all lymph nodes that were analysed, with the most striking effects in the PLNs (the brachial and inguinal lymph nodes) and the mesenteric lymph nodes (Fig. 1b). By contrast, the number of leukocytes and lymphocytes in

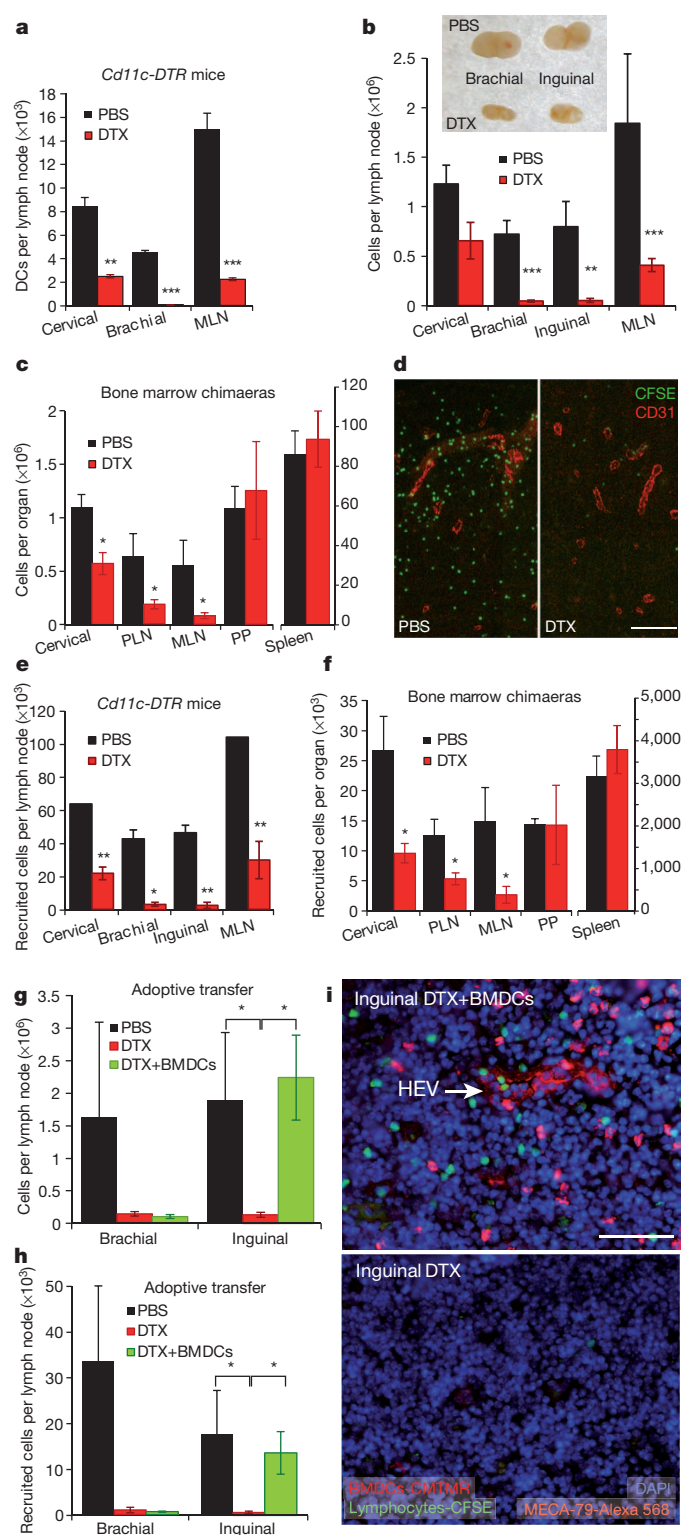
the blood was not significantly altered by DC depletion (Supplementary Fig. 2). We observed weight loss after repeated DTX treatment of *Cd11c-DTR* mice (Supplementary Fig. 3a), as previously reported<sup>16,17</sup>. These deleterious effects of DTX are thought to result from the expression of the DTR on non-haematopoietic radioresistant cells<sup>17</sup>. To confirm our initial observations, we therefore generated bone marrow chimaeras in which *Cd11c-DTR* expression was restricted to the haematopoietic compartment<sup>17</sup>. Lethally irradiated wild-type mice were reconstituted with bone marrow from *Cd11c-DTR* transgenic mice, and 3 months later, they were treated with DTX for 10 days, without any adverse side effects (Supplementary Fig. 3a). The total cellularity in the lymph nodes was reduced in DTX-treated chimaeric mice, whereas the cellularity in the Peyer's patches and the spleen was not affected (Fig. 1c). By contrast, we observed no significant changes in the cellularity of the lymph nodes after prolonged DTX treatment of lethally irradiated *Cd11c-DTR* transgenic mice that had been reconstituted with wild-type bone marrow (Supplementary Fig. 3b, c), despite these mice rapidly succumbing to weight loss and death, as previously described<sup>17</sup>.

To determine whether the reduced cellularity in the absence of DCs results from inhibition of lymphocyte entry to lymph nodes, we next performed short-term homing assays. The homing of lymphocytes from the blood to the PLNs and mucosal lymph nodes was strongly reduced in DTX-treated *Cd11c-DTR* transgenic and chimaeric mice compared with PBS-treated mice (Fig. 1d–f). The effects of DC depletion on lymphocyte homing and lymph node cellularity in *Cd11c-DTR* chimaeric mice were completely reversible after DTX injections were stopped (Supplementary Fig. 4). In addition, we found that adoptive transfer of wild-type, bone-marrow-derived CD11c<sup>+</sup> DCs (wild-type BMDCs; Supplementary Fig. 5) to DTX-treated *Cd11c-DTR* transgenic mice could recover the normal cellularity of, and the homing of naive lymphocytes to, the draining lymph nodes (Fig. 1g, h). The adoptively transferred CD11c<sup>+</sup> BMDCs were located around HEVs expressing antigens recognized by MECA-79 (denoted MECA-79<sup>+</sup> HEVs) in the draining lymph nodes (Fig. 1i), and *Ccr7*<sup>−/−</sup> BMDCs had a lower capacity than wild-type BMDCs to recover lymph node cellularity and homing to lymph nodes (Supplementary Fig. 5). Together, these results indicate that CD11c<sup>+</sup> DCs are essential for lymphocyte homing to lymph nodes (Supplementary Discussion).

We next analysed the HEV phenotype and observed a marked down-regulation of HEV-specific markers (FucT-VII, GlcNAc6ST-2 and the L-selectin counter receptor GLYCAM1 (ref. 6), as well as MECA-79 antigens) in the PLNs of DC-depleted mice (Fig. 2a and Supplementary Fig. 6). By contrast, the expression of pan-endothelial cell markers, such as CD31 (also known as PECAM1), was unaffected (Fig. 2a). Strikingly, the expression of MADCAM1, a marker of immature neonatal HEVs<sup>18</sup>, was induced on the blood vessels in the inguinal lymph nodes of DC-depleted mice (Fig. 2a). Thus, *in vivo* depletion of CD11c<sup>+</sup> DCs induces specific changes in HEV phenotype, with a reversion from the mature adult HEV phenotype (MECA-79<sup>hi</sup>GLYCAM1<sup>hi</sup>FucT-VII<sup>hi</sup>GlcNAc6ST-2<sup>hi</sup>MADCAM1<sup>lo</sup>) to the immature HEV phenotype

<sup>1</sup>CNRS, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, F-31077 Toulouse, France. <sup>2</sup>Université de Toulouse, UPS, Institut de Pharmacologie et de Biologie Structurale, F-31077 Toulouse, France.





(MECA-79<sup>lo</sup>GLYCAM1<sup>lo</sup>FucT-VII<sup>lo</sup>GlcNAc6ST-2<sup>lo</sup>MADCAM1<sup>hi</sup>), which is found in PLNs at birth<sup>13,14,18</sup> or after the occlusion of the afferent lymphatics<sup>12–14</sup>.

To determine whether the changes in HEV phenotype occur at the RNA level, we purified MECA-79<sup>+</sup>CD31<sup>+</sup> HEV and MECA-79<sup>−</sup>CD31<sup>+</sup> non-HEV endothelial cells from the PLNs of *Cdl1c-DTR* transgenic mice that had been treated with DTX or PBS (Supplementary Fig. 7). Although the percentage of MECA-79<sup>+</sup> cells in the CD31<sup>+</sup> endothelial cell population was halved after DTX treatment, and the intensity of MECA-79 staining on the remaining

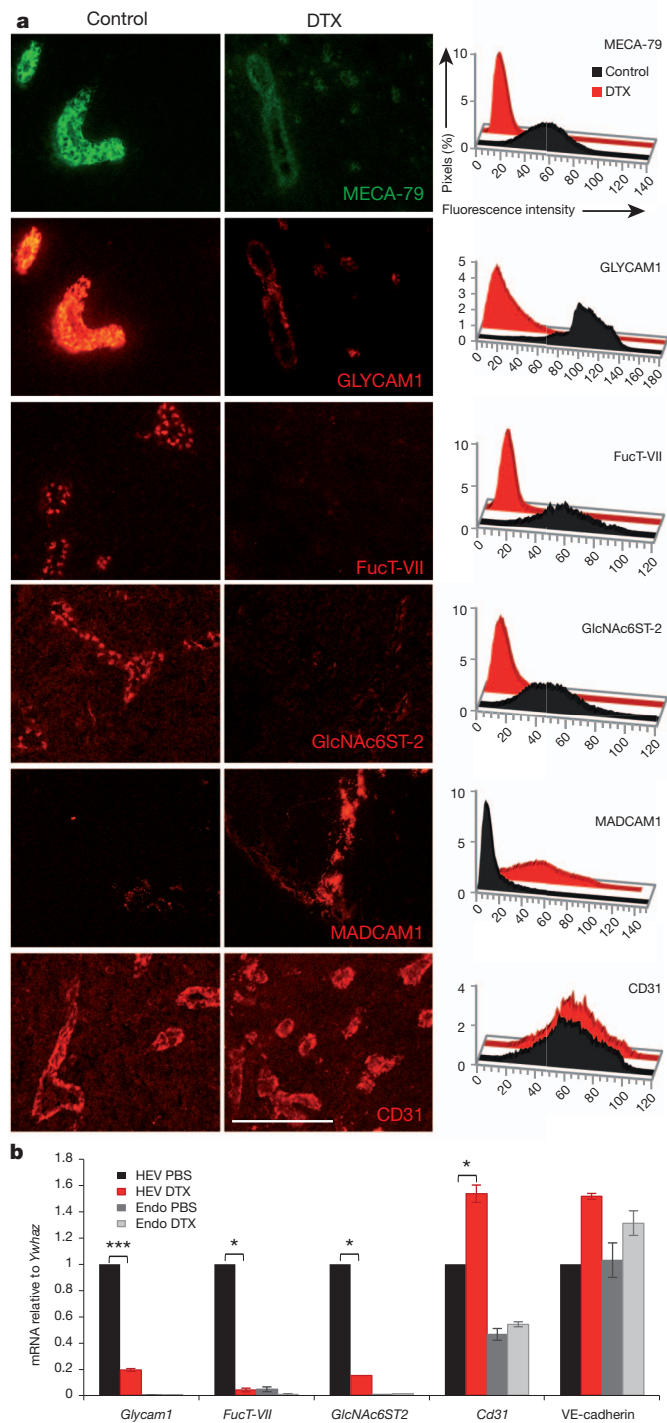
**Figure 1** | CD11c<sup>+</sup> DCs are required for lymphocyte homing to lymph nodes. **a–f**, Depletion of CD11c<sup>+</sup> DCs in *Cdl1c-DTR* transgenic mice (**a, b, d, e**) or in *Cdl1c-DTR* bone marrow chimaeras (**c, f**) reduces DC numbers (**a**), lymph node cellularity (**b, c**) and lymphocyte homing (**d–f**) in PLNs (the brachial and inguinal lymph nodes) and mucosal lymph nodes (the cervical and mesenteric lymph nodes (MLNs)). Lymphocyte homing was examined by immunofluorescence staining (**d**, shown is a PLN section with lymphocytes labelled with CFSE and blood vessels labelled with CD31) and by flow cytometry (**e, f**). (**a, b**,  $n = 7–11$ , four experiments; **c, f**,  $n = 4–8$ , three experiments; **d, e**,  $n = 4$ , two experiments.) **g–i**, Adoptive transfer of BMDCs restores cellularity (**g**) and homing (**h, i**) to the draining (inguinal) lymph nodes of DTX-treated *Cdl1c-DTR* transgenic mice ( $n = 4$  per group). **a–c, e–h**, Error bars, s.e.m.; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . **d, i**, Scale bar, 100  $\mu$ m. DAPI, 4',6-diamidino-2-phenylindole; PP, Peyer's patch.

MECA-79<sup>+</sup> cells was also reduced (by  $\sim 75\%$ ), it was nevertheless possible to isolate these cells by flow cytometry. Analysis of HEV-specific gene expression by reverse transcription followed by quantitative PCR (RT-qPCR) showed that expression of the HEV-specific genes *Glycam1*, *FucT-VII* (also known as *Fut7*) and *GlcNAc6ST2* (also known as *Chst4*) was strongly reduced after DTX treatment, whereas the expression of the genes encoding the pan-endothelial cell markers CD31 and VE-cadherin was not downregulated after depletion of DCs (Fig. 2b). The expression of several other genes involved in the multi-step lymphocyte adhesion cascade (*Ccl21*, *Icam1*, *Cxcl13* and *Vcam1*) was also not altered in DC-depleted mice (Supplementary Fig. 8). Therefore, CD11c<sup>+</sup> DCs are essential for the maintenance of HEV-specific gene expression.

We then analysed lymphocyte–HEV interactions in the inguinal lymph nodes from *Cdl1c-DTR* chimaeric mice by using intravital microscopy. As previously described<sup>19</sup>, the HEV phenotype is found only in venular orders III to V (Fig. 3a), which support the bulk of lymphocyte trafficking to lymph nodes. The rolling fraction of wild-type lymphocytes in venular orders II to V was not significantly different between *Cdl1c-DTR* chimaeric mice that had been treated with DTX and those treated with PBS (Fig. 3b). However, the fraction of lymphocytes that transitioned from primary rolling to secondary firm adhesion (the sticking fraction) was markedly decreased in HEVs (orders III to V) from DC-depleted mice (Fig. 3c). In addition, the rolling velocity of lymphocytes in HEVs was significantly increased (Fig. 3d, e). For instance, the median rolling velocity of lymphocytes in order IV HEVs increased from  $20 \mu\text{m s}^{-1}$  in PBS-treated mice to  $92 \mu\text{m s}^{-1}$  in DTX-treated mice. Thus, *in vivo* depletion of CD11c<sup>+</sup> DCs in *Cdl1c-DTR* chimaeric mice elicited considerable alterations in lymphocyte rolling velocity and lymphocyte sticking inside lymph node HEVs.

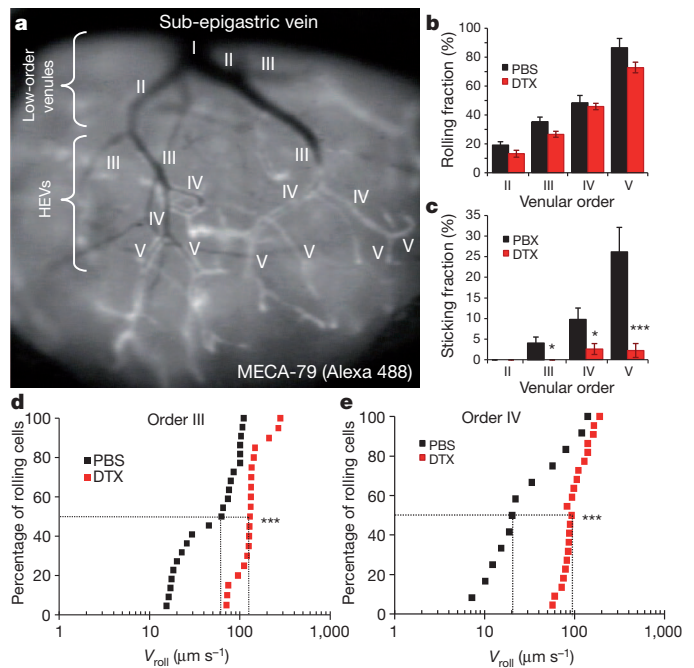
An important question raised by our observations is whether CD11c<sup>+</sup> DCs regulate HEV phenotype directly or through other cell types that are found in the lymph nodes. Analyses in *Rag2*<sup>−/−</sup> mice and *Rag2*<sup>−/−</sup>*Cdl1c-DTR* transgenic mice showed that lymphocytes are not essential for the homeostatic maintenance of HEV phenotype and that the effects of CD11c<sup>+</sup> DCs on lymphocyte homing occur in the absence of T and B lymphocytes (Supplementary Fig. 9). Therefore, we next considered the possibility that DCs, which are strategically positioned close to HEV walls *in vivo*<sup>20,21</sup>, may regulate HEV phenotype directly. We developed a co-culture model of MECA-79<sup>+</sup>CD31<sup>+</sup> HEV endothelial cells with CD11c<sup>+</sup>MHC class II<sup>+</sup> DCs (Fig. 4a), both of which were isolated from the PLNs of wild-type mice (Supplementary Fig. 10).

Setting up the culture conditions for the purified HEV endothelial cells turned out to be technically challenging because these cells were available in a limited quantity (constituting 0.02% of total lymph node cells) and nothing was known about the conditions required to maintain the cells *ex vivo*. We eventually found that culture of the purified HEV endothelial cells on plates coated with fibronectin and type I collagen in an appropriate endothelial/DC culture medium allowed the survival of the cells *ex vivo* for at least 12 days. However, HEV-specific gene



**Figure 2 | CD11c<sup>+</sup> DCs are essential for the homeostatic maintenance of HEV phenotype.** **a**, PLN sections from control mice (control) or *Cd11c-DTR* transgenic mice (DTX) treated with DTX for 8 days were stained with MECA-79 or antibodies specific for the indicated proteins (left). Quantitative image analysis of the immunofluorescence staining is also shown (right). Scale bar, 100  $\mu$ m. **b**, qPCR analysis was performed using total RNA from HEV endothelial cells (HEV) and non-HEV endothelial cells (Endo) that were sorted by flow cytometry after isolation from DTX-treated or PBS-treated *Cd11c-DTR* transgenic mice. The mean and s.d. from triplicate qPCR runs are plotted. Data are representative of three independent experiments. \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ .

expression (as determined by measuring the levels of *Glycam1* messenger RNA) was strongly reduced when HEV endothelial cells were grown alone *ex vivo* (Fig. 4b). Strikingly, when MECA-79<sup>+</sup> HEV endothelial cells were co-cultured with CD11c<sup>+</sup> DCs purified from

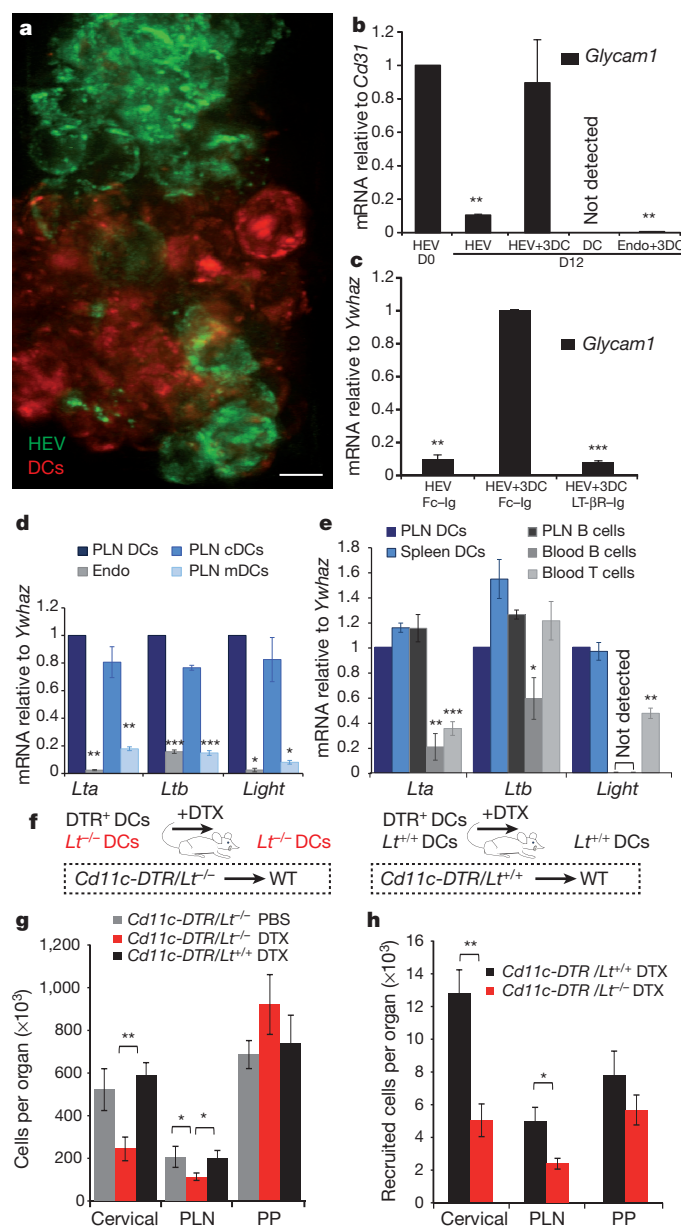


**Figure 3 | CD11c<sup>+</sup> DCs control the rolling velocity and firm adhesion of lymphocytes inside PLN HEVs.** **a**, Visualization of the inguinal lymph node microcirculation by intravital microscopy. **b–e**, The rolling fraction (**b**), sticking fraction (**c**) and cumulative rolling velocities (**d**, **e**) of fluorescently labelled wild-type lymphocytes were determined in the indicated venules. Data were collected from three DTX- and three PBS-treated *Cd11c-DTR* bone marrow chimaeras. **d**, **e**, Data represent the percentages of cells rolling at or below a given velocity ( $V_{roll}$ ). The median  $V_{roll}$  in order III venules was 62.5  $\mu$ m s<sup>-1</sup> in PBS-treated chimaeras and 129.6  $\mu$ m s<sup>-1</sup> in DTX-treated chimaeras (**d**). The median  $V_{roll}$  in order IV venules was 20.3  $\mu$ m s<sup>-1</sup> in PBS-treated chimaeras and 92.9  $\mu$ m s<sup>-1</sup> in DTX-treated chimaeras (**e**). (**d**, PBS,  $n = 22$  cells, 5 venules and 3 mice; DTX,  $n = 20$  cells, 6 venules and 3 mice; **e**, PBS,  $n = 12$  cells, 4 venules and 3 mice; DTX,  $n = 22$  cells, 10 venules and 3 mice.) Error bars, s.e.m. \*,  $P < 0.05$ ; \*\*\*,  $P < 0.001$ .

wild-type mice, *Glycam1* mRNA levels were maintained at levels similar to those found in HEVs that were freshly purified from the PLNs (Fig. 4b and Supplementary Fig. 11), indicating that CD11c<sup>+</sup> DCs regulate HEV phenotype directly. *FucT-VII* and *GlcNAc6ST-2* mRNA levels could not be quantified in these experiments, probably because our qPCR assay was not sensitive enough or because all of the properties of HEVs may not be maintained in the co-cultures.

The lymphotoxin- $\beta$  receptor (LT- $\beta$ R), which is expressed on HEV endothelial cells<sup>22</sup>, has been shown to have a crucial role in the homeostatic maintenance of HEVs<sup>23</sup>, so we tested the possibility that this receptor may be involved in the cross-talk between HEV endothelial cells and CD11c<sup>+</sup> DCs. We found that interrupting LT- $\beta$ R signalling through the addition of LT- $\beta$ R-immunoglobulin soluble inhibitor (but not control Fc-immunoglobulin) to the co-cultures abrogated the enhancing effect of CD11c<sup>+</sup> DCs on *Glycam1* expression *ex vivo* (Fig. 4c and Supplementary Fig. 11). CD11c<sup>+</sup> DCs isolated from the PLNs (Supplementary Fig. 10) or the spleen expressed transcripts encoding the LT- $\beta$ R ligands lymphotoxin- $\alpha$  (LT- $\alpha$ ), LT- $\beta$  and LIGHT (Fig. 4d, e), and the expression levels of these ligands were higher in CD11c<sup>hi</sup>MHC class II<sup>med</sup> DCs (classical DCs (cDCs); the major population depleted in *Cd11c-DTR* transgenic mice) than in CD11c<sup>med</sup>MHC class II<sup>hi</sup> DCs (migratory DCs (mDCs)) (Fig. 4d). Interestingly, the expression levels of LT- $\alpha$ - and LT- $\beta$ -encoding transcripts (*Lta* and *Ltb* mRNA) in DCs were similar to those of naive B cells from PLNs (Fig. 4e), and such levels are sufficient for the maintenance of follicular DC networks in the lymph nodes<sup>24</sup>. The expression levels of *Lta* and *Ltb* in DCs may thus be appropriate for homeostatic maintenance of HEVs.





**Figure 4** | DC-derived LT is important for lymphocyte homing to lymph nodes *in vivo*. **a**, Immunofluorescence staining of HEVs (MECA-79 staining) co-cultured with DCs. Scale bar, 10  $\mu$ m. **b–e**, qPCR analysis of *Glycam1* (**b**, **c**), *Lta*, *Ltb* and *Light* (**d**, **e**) mRNA levels. Mean and s.d. from triplicate qPCR runs are plotted (**b–e**). D, day; Endo, non-HEV endothelial cells; HEV+3DC (Endo+3DC), the number of DCs in the co-cultures was threefold the number of HEV (Endo) endothelial cells. **f–h**, Analysis of *Cd11c-DTR/Lt<sup>-/-</sup>* and *Cd11c-DTR/Lt<sup>+/+</sup>* chimaeric mice (**f**). Lethally irradiated wild-type (WT) mice were reconstituted with the indicated bone marrow mixture (dashed boxes). Cellularity in (**g**) and lymphocyte homing to (**h**) lymph nodes were determined by flow cytometry after 10 days of DTX treatment ( $n = 4$  mice per group). Data are representative of two independent experiments. **b–e**, **g**, **h**, Error bars, s.e.m. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .

To examine the potential role of DC-derived LT *in vivo*, we generated mixed bone marrow chimaeras by reconstituting lethally irradiated wild-type mice with 50% bone marrow from *Cd11c-DTR* transgenic mice and 50% bone marrow from *Lta*-deficient (denoted *Lt<sup>-/-</sup>*) or wild-type (*Lt<sup>+/+</sup>*) mice (Fig. 4f). Injection of DTX into the *Cd11c-DTR/Lt<sup>-/-</sup>* or *Cd11c-DTR/Lt<sup>+/+</sup>* chimaeric mice induced depletion of *Lt<sup>+/+</sup>* DCs harbouring the *Cd11c-DTR* transgene but spared non-transgenic *Lt<sup>-/-</sup>* and *Lt<sup>+/+</sup>* DCs (Supplementary Fig. 12). The cellularity of the PLNs and cervical lymph nodes and lymphocyte homing to these lymph nodes (Fig. 4g, h), but not to the Peyer's patches

and the spleen (Supplementary Fig. 12), were significantly reduced in DTX-treated *Cd11c-DTR/Lt<sup>-/-</sup>* chimaeric mice (which retained only *Lt*-deficient CD11c<sup>+</sup> DCs) compared with DTX-treated *Cd11c-DTR/Lt<sup>+/+</sup>* chimaeric mice. DC-derived LT is thus important *in vivo* for HEV-mediated lymphocyte recruitment to lymph nodes.

In conclusion, our findings uncover a previously unsuspected and important function of DCs in the control of HEV phenotype and lymphocyte entry to lymph nodes during homeostasis in adult mice. DCs thus have a dual role; they function as antigen-presenting cells<sup>3</sup>, but they also regulate lymphocyte recirculation to ensure that the appropriate lymphocytes will come into contact with their cognate antigens. The role of DCs in the maintenance of blood vessel phenotype is reminiscent of the role of astrocytes in the brain<sup>4,25</sup>. However, astrocytes induce blood–brain barrier characteristics that limit the entry of molecules and cells (including lymphocytes) to the brain, whereas DCs regulate HEV characteristics that allow lymphocyte entry to lymph nodes. The development of HEVs is observed in many chronic inflammatory diseases<sup>4</sup>, and CD11c<sup>+</sup> DCs from inflamed tissues produce LT<sup>26,27</sup> and regulate HEV angiogenesis<sup>28</sup>. A better understanding of the cross-talk between DCs and HEVs may thus provide a basis for novel approaches to control the development of HEVs in clinical situations.

## METHODS SUMMARY

*Cd11c-DTR* mice (B6.FVB-Tg<sup>ltagax-DTR/EGFP.57</sup>Lan/J) were purchased from the EMMA European network, and C57BL/6 mice were purchased from Charles River Laboratories. C57BL/6:*Lt<sup>-/-</sup>* mice were provided by H. Korner and S. A. Nedospasov. For bone marrow chimaeras, a single-cell suspension generated from the flushed bone marrow of femurs and tibias was injected intravenously into mice that had been irradiated with 900 rad. Chimaeras were incubated for at least 12 weeks before use. All mice were handled according to institutional guidelines under protocols approved by the Institut de Pharmacologie et de Biologie Structurale and Région Midi-Pyrénées animal care committees. For DC depletion, mice were treated with DTX (Sigma), which was administered intraperitoneally at 6 ng per gram (body weight) every two days. BMDCs were generated in medium containing mouse GM-CSF, as described previously<sup>29</sup>, and then treated with lipopolysaccharide and injected subcutaneously into recipient mice every 2 days. Homing assays and intravital microscopy analyses with fluorescent lymphocytes were performed as previously described<sup>19,20</sup>. Immunofluorescence staining was performed on Bouin-fixed, paraffin-embedded or cryopreserved mouse lymph node sections. HEV endothelial cells (CD45<sup>+</sup>CD31<sup>+</sup>MECA-79<sup>+</sup>) and non-HEV endothelial cells (CD45<sup>+</sup>CD31<sup>+</sup>MECA-79<sup>-</sup>) were isolated from lymph node stromal-cell suspensions (using pooled PLNs from 5–20 mice) by flow cytometry using a FACSaria II cell sorter. CD11c<sup>+</sup>MHC class II<sup>+</sup> wild-type DCs were sorted from non-adherent cells in the same pool of PLN cells. For co-culture assays, sorted cells were seeded in 96-well tissue culture plates coated with rat fibronectin, type I collagen and gelatine (Sigma) and were grown *ex vivo* for 11–12 days in endothelial/DC culture medium. LT- $\beta$ R-immunoglobulin fusion protein (5  $\mu$ g ml<sup>-1</sup>; R&D Systems) or control human Fc-immunoglobulin was added to the co-cultures every 2 days. qPCR was performed as described previously<sup>30</sup>. Statistical analysis was performed using the Mann–Whitney test (for lymph node cellularity and velocity histograms) or an unpaired Student's *t*-test (for homing assays, and rolling and sticking fractions). Differences were considered statistically significant when  $P < 0.05$ .

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 July; accepted 2 September 2011.

Published online 13 November 2011.

- Butcher, E. C. & Picker, L. J. Lymphocyte homing and homeostasis. *Science* **272**, 60–66 (1996).
- von Andrian, U. H. & Mempel, T. R. Homing and cellular traffic in lymph nodes. *Nature Rev. Immunol.* **3**, 867–878 (2003).
- Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
- Girard, J. P. & Springer, T. A. High endothelial venules (HEVs): specialized endothelium for lymphocyte migration. *Immunol. Today* **16**, 449–457 (1995).
- Miyasaka, M. & Tanaka, T. Lymphocyte trafficking across high endothelial venules: dogmas and enigmas. *Nature Rev. Immunol.* **4**, 360–370 (2004).
- Rosen, S. D. Ligands for L-selectin: homing, inflammation, and beyond. *Annu. Rev. Immunol.* **22**, 129–156 (2004).



7. Streeter, P. R., Rouse, B. T. & Butcher, E. C. Immunohistologic and functional characterization of a vascular addressin involved in lymphocyte homing into peripheral lymph nodes. *J. Cell Biol.* **107**, 1853–1862 (1988).
8. Uchimura, K. *et al.* A major class of L-selectin ligands is eliminated in mice deficient in two sulfotransferases expressed in high endothelial venules. *Nature Immunol.* **6**, 1105–1113 (2005).
9. Kawashima, H. *et al.* N-acetylglucosamine-6-O-sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial venules. *Nature Immunol.* **6**, 1096–1104 (2005).
10. Maly, P. *et al.* The  $\alpha(1,3)$ fucosyltransferase Fuc-TVII controls leukocyte trafficking through an essential role in L-, E-, and P-selectin ligand biosynthesis. *Cell* **86**, 643–653 (1996).
11. Hendriks, H. R. & Eestermans, I. L. Disappearance and reappearance of high endothelial venules and immigrating lymphocytes in lymph nodes deprived of afferent lymphatic vessels: a possible regulatory role of macrophages in lymphocyte migration. *Eur. J. Immunol.* **13**, 663–669 (1983).
12. Mebius, R. E., Streeter, P. R., Breve, J., Duijvestijn, A. M. & Kraal, G. The influence of afferent lymphatic vessel interruption on vascular addressin expression. *J. Cell Biol.* **115**, 85–95 (1991).
13. Swarte, V. V. *et al.* Regulation of fucosyltransferase-VII expression in peripheral lymph node high endothelial venules. *Eur. J. Immunol.* **28**, 3040–3047 (1998).
14. Mebius, R. E. *et al.* Expression of GlyCAM-1, an endothelial ligand for L-selectin, is affected by afferent lymphatic flow. *J. Immunol.* **151**, 6769–6776 (1993).
15. Lacorre, D. A. *et al.* Plasticity of endothelial cells: rapid dedifferentiation of freshly isolated high endothelial venule endothelial cells outside the lymphoid tissue microenvironment. *Blood* **103**, 4164–4172 (2004).
16. Jung, S. *et al.* *In vivo* depletion of CD11c<sup>+</sup> dendritic cells abrogates priming of CD8<sup>+</sup> T cells by exogenous cell-associated antigens. *Immunity* **17**, 211–220 (2002).
17. Zaft, T., Sapozhnikov, A., Krauthgamer, R., Littman, D. R. & Jung, S. CD11c<sup>high</sup> dendritic cell ablation impairs lymphopenia-driven proliferation of naive and memory CD8<sup>+</sup> T cells. *J. Immunol.* **175**, 6428–6435 (2005).
18. Mebius, R. E., Streeter, P. R., Michie, S., Butcher, E. C. & Weissman, I. L. A developmental switch in lymphocyte homing receptor and endothelial vascular addressin expression regulates lymphocyte homing and permits CD4<sup>+</sup> CD3<sup>+</sup> cells to colonize lymph nodes. *Proc. Natl Acad. Sci. USA* **93**, 11019–11024 (1996).
19. Von Andrian, U. H. Intravital microscopy of the peripheral lymph node microcirculation in mice. *Microcirculation* **3**, 287–300 (1996).
20. Bajenoff, M., Granjeaud, S. & Guerder, S. The strategy of T cell antigen-presenting cell encounter in antigen-draining lymph nodes revealed by imaging of initial T cell activation. *J. Exp. Med.* **198**, 715–724 (2003).
21. Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154–159 (2004).
22. Liao, S. & Ruddle, N. H. Synchrony of high endothelial venules and lymphatic vessels revealed by immunization. *J. Immunol.* **177**, 3369–3379 (2006).
23. Browning, J. L. *et al.* Lymphotoxin- $\beta$  receptor signaling is required for the homeostatic control of HEV differentiation and function. *Immunity* **23**, 539–550 (2005).
24. Ansel, K. M. *et al.* A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature* **406**, 309–314 (2000).
25. Janzer, R. C. & Raff, M. C. Astrocytes induce blood–brain barrier properties in endothelial cells. *Nature* **325**, 253–257 (1987).
26. GeurtsvanKessel, C. H. *et al.* Dendritic cells are crucial for maintenance of tertiary lymphoid structures in the lung of influenza virus-infected mice. *J. Exp. Med.* **206**, 2339–2349 (2009).
27. Muniz, L. R., Pacer, M. E., Lira, S. A. & Furtado, G. C. A critical role for dendritic cells in the formation of lymphatic vessels within tertiary lymphoid structures. *J. Immunol.* **187**, 828–834 (2011).
28. Webster, B. *et al.* Regulation of lymph node vascular growth by dendritic cells. *J. Exp. Med.* **203**, 1903–1913 (2006).
29. Lutz, M. B. *et al.* An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. *J. Immunol. Methods* **223**, 77–92 (1999).
30. Cayrol, C. *et al.* The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. *Blood* **109**, 584–594 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to S. Rosen, J. Lowe and J. Browning for the gift of antibodies, to H. Korner, T. Winkler, E. Donnadiou, M. Lipp, O. Lantz, B. Ryffel, L. Brault, S. A. Nedospasov and S. Jung for providing mice, to J. van Meerwijk for critical reading of the manuscript, and to N. Ortega, E. Bellard and F. El-Fahqui-Olive for help with immunohistochemistry, intravital microscopy and cell sorting, respectively. We thank the IPBS TRI imaging facility, A. Dujardin and the IPBS animal and transgenic facilities for help with animal experiments, and F. Viala for preparation of the figures. This work was supported by grants from the Ligue Nationale Contre le Cancer (Equipe labellisée Ligue 2009 to J.-P.G.) and the Association pour la Recherche contre le Cancer (ARC, equipment 8505). C.M. was supported by fellowships from the French Ministry of Research, the ARC and Fondation RITC.

**Author Contributions** C.M. designed the study, performed all of the experiments and analysed the data. J.-P.G. designed the study, analysed the data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.-P.G. (Jean-Philippe.Girard@ipbs.fr).

## METHODS

**Mice.** C57BL/6 wild type mice were purchased from Charles River Laboratories, and *Cd11c-DTR* transgenic mice (B6.FVB-Tg<sup>lga-x-DTR/EGFP.57</sup>/Lan/J)<sup>16</sup> were purchased from the EMMA European network. *Cd11c-DTR* mice were maintained on the C57BL/6 and C57BL/6:Rag2<sup>-/-</sup> genetic backgrounds. All *CD11c-DTR* transgenic mice used in this study were heterozygous transgenic. C57BL/6:Lf<sup>-/-</sup> mice<sup>31,32</sup> were provided by H. Körner and S. A. Nedospasov. C57BL/6:Cd45.1 and C57BL/6:Rag2<sup>-/-</sup> mice were obtained from O. Lantz, and C57BL/6:CCR7<sup>-/-</sup> mice were obtained from M. Lipp and E. Donnadieu. All mice were bred under specific-pathogen-free conditions and handled according to institutional guidelines under protocols approved by the Institut de Pharmacologie et de Biologie Structurale and Région Midi-Pyrénées animal care committees.

**Bone marrow chimaeras.** A single-cell suspension generated from the flushed bone marrow of femurs and tibias from donor mice was injected intravenously into mice that had been irradiated with 900 rad. Chimaeras were incubated for at least 12 weeks before use. Mixed bone marrow chimaeras<sup>33</sup> were generated by transferring a mixture of 50% bone marrow from *Cd11c-DTR* transgenic mice and 50% bone marrow from *Lt<sup>+/+</sup>* or *Lt<sup>-/-</sup>* mice into lethally irradiated wild-type recipient mice.

**Depletion of CD11c<sup>+</sup> DCs.** For systemic DC depletion, mice were given DTX (Sigma) intraperitoneally at a dose of 6 ng per gram (body weight) every two days. As controls, mice received intraperitoneal injections of PBS vehicle. *Cd11c-DTR* transgenic and chimaeric mice were treated with DTX or PBS for 8 and 10 days, respectively. To evaluate the efficacy of DC depletion, spleens and lymph nodes were analysed by flow cytometry for the presence of CD11c<sup>+</sup>GFP<sup>+</sup> cells or CD11c<sup>+</sup>MHC class II<sup>+</sup> cells.

**Analysis of blood and lymph nodes.** Peripheral blood was obtained by cardiac puncture and was stored in EDTA-containing tubes at 4 °C until analysis with an automated haematological analyser (Sysmex XT 2000i). Freshly isolated mucosal lymph nodes (three cervical and one mesenteric) and PLNs (one inguinal and one brachial) were digested with type D collagenase (Roche laboratories), and single-cell suspensions were counted (size >7 µm) with a Z1 Particle Counter (Beckman Coulter).

**Analysis of surface markers.** The following monoclonal antibodies for cell surface staining were purchased from BD Biosciences: anti-CD11c (HL3), anti-CD45.1 (A20), anti-CD45.2 (104), anti-Gr1 (RB6-8C5), anti-CD3 molecular complex (17A2) and anti-CD19 (1D3) antibodies. The following monoclonal antibodies were purchased from eBioscience: anti-CD11c (N418) and anti-MHC class II (I-A/I-E) (M5/114.15.2) antibodies. After a 20-min blocking step on ice (in PBS containing 1% FCS, 5% normal mouse serum, 5 mM EDTA, 0.1% NaN<sub>3</sub>, 5 µg ml<sup>-1</sup> anti-CD16/CD32 (2.4G2, BD Biosciences)), cells were incubated with conjugated antibodies (conjugates were FITC, Alexa 488, PE, PE-Cy7, Alexa 700, biotin or APC) diluted in FACS buffer (PBS containing 1% FCS, 5 mM EDTA and 0.1% NaN<sub>3</sub>). When biotinylated monoclonal antibodies were used, cells were incubated with APC- or PE-Cy7-conjugated streptavidin (BD Biosciences). Flow cytometry analyses were performed on a FACSCalibur or an LSRII flow cytometer (BD Biosciences).

**Immunofluorescence staining.** Immunofluorescence staining was performed on 5 µm sections from Bouin-fixed, paraffin-embedded or cryopreserved mouse lymph nodes, as described previously<sup>34</sup>, using the following primary antibodies: the rat monoclonal antibodies MECA-79 (ATCC) and MECA-367 (anti-MADCAM1, Pharmingen; for cryosections); rabbit polyclonal antibodies against CD31 (Abcam), GLYCAM1 (CAM02; provided by S. D. Rosen), GlcNAc6ST-2 (GST-3, provided by S. D. Rosen) and FucT-VII (provided by J. B. Lowe). For quantitative image analysis of the immunofluorescence staining, lymph node sections were immunostained in parallel using the same dilutions of antibodies and reagents, and fluorescent images, which were captured using identical exposure times and settings, were analysed with ImageJ software, as described previously<sup>35</sup>.

**In vivo homing assays.** A single-cell suspension of naive lymphocytes was prepared from a pool of spleen (after lysis of the red blood cells in ammonium chloride, potassium bicarbonate and EDTA) and PLNs from C57BL/6 mice. Lymphocytes labelled with CFSE (Molecular Probes, Invitrogen) were injected intravenously into recipient mice. The number of fluorescent cells recruited to each lymph node was determined 4 h after injection, by flow cytometry or by fluorescence microscopy, as described previously<sup>20</sup>.

**Adoptive transfer of BMDCs.** BMDCs were generated in medium containing mouse GM-CSF, as described previously<sup>29</sup>. BMDCs were then incubated with 0.5 µg ml<sup>-1</sup> lipopolysaccharide (Sigma-Aldrich) for 1 h at 37 °C, washed, labelled with CMTMR (Molecular Probes, Invitrogen) and injected (5 × 10<sup>6</sup> cells per mouse) into two subcutaneous sites (the footpad and thigh). Cells were injected every 2 days at the time of intraperitoneal DTX injection, for 8 days (a total of 2 × 10<sup>7</sup> BMDCs per mouse). In some experiments, BMDCs were prepared in the absence of lipopolysaccharide (Supplementary Fig. 5d, e).

**Intravital microscopy.** Intravital microscopy was performed as described previously<sup>19,36,37</sup>. Naive lymphocytes labelled with calcein (Molecular Probes, Invitrogen) were injected into the right femoral artery, and cell behaviour in lymph node venules was assessed as described previously<sup>19</sup>. To determine the rolling fractions (the percentage of rolling cells in the total flux of cells in each venule) and the sticking fractions (the percentage of rolling cells that subsequently arrested for >30 s), the data were generated from three DTX-treated *Cd11c-DTR* chimaeric mice (number of venules analysed/venular order: 2/II, 4/III, 8/IV and 16/V) and three PBS-treated *Cd11c-DTR* chimaeric mice (number of venules analysed/venular order: 2/II, 3/III, 5/IV and 7/V) (with DTX and PBS treatment for 10 days). In some experiments, Alexa-488-conjugated MECA-79 monoclonal antibody (Molecular probes A488 labelling kit) was injected intravenously to visualize the HEV network.

**Isolation of HEV endothelial cells and CD11c<sup>+</sup>MHC class II<sup>+</sup> DCs from PLNs.** PLNs were pooled and gently squeezed, and the non-adherent cells were washed out. The stromal elements were digested with type II collagenase (Gibco) for 1 h at 37 °C, and the single-cell suspensions obtained were passed through a 40 µm cell strainer (BD Biosciences) and stained with anti-CD45 (30-F11, BD Biosciences), anti-CD31 (MEC13.3, BD Biosciences) and MECA-79 monoclonal antibodies for 1 h at 4 °C. In some experiments, one round of CD45<sup>+</sup> cell depletion with Dynabeads was performed. HEV endothelial cells (CD45<sup>+</sup>CD31<sup>+</sup>MECA-79<sup>+</sup>) and non-HEV endothelial cells (CD45<sup>+</sup>CD31<sup>+</sup>MECA-79<sup>-</sup>) were isolated by cell sorting using a FACSAria II cell sorter (BD Biosciences). CD11c<sup>+</sup>MHC class II<sup>+</sup> wild-type DCs, CD11c<sup>hi</sup>MHC class II<sup>med</sup> DCs (classical DCs (cDCs)) and CD11c<sup>med</sup>MHC class II<sup>hi</sup> DCs (migratory DCs (mDCs)) were sorted from non-adherent cells from the same PLNs. Anti-rat/hamster Igk chain compensation particles (BD CompBeads, BD Biosciences), single stained with each of the antibodies, were used as compensation controls.

**Co-culture assays.** Sorted cells were seeded in 96-well tissue culture plates coated with rat fibronectin (Sigma) and type I collagen (Sigma) in 0.2% gelatine (Sigma). Cells were grown *ex vivo* for 11–12 days in endothelial/DC culture medium: a mixture of 50% BD Endothelial Cell Culture Medium and 50% RPMI-1640 supplemented with 1 × Endothelial Cell Growth Supplement (ECGS, BD Biosciences), 1 × epidermal growth factor (EGF, BD Biosciences), 15% FCS, 8% mouse serum, L-glutamine, non-essential amino acids, penicillin/streptomycin (Invitrogen), 10 ng ml<sup>-1</sup> recombinant mouse GM-CSF (tebu-bio) and 2-mercaptoethanol (Sigma). Survival of the isolated HEV endothelial cells in culture was sensitive to the batches of ECGS and serum; in some experiments, 1.5 × ECGS, 1.5 × EGF, 20% FCS, 4% newborn calf serum and 4% KnockOut Serum Replacement (Invitrogen) were used to increase the survival of the cells. In some experiments, LT-βR-immunoglobulin fusion protein (5 µg ml<sup>-1</sup>, R&D Systems) or control human Fc-immunoglobulin (5 µg ml<sup>-1</sup>, Chemicon) was added to the co-cultures every 2 days.

**qPCR.** Total RNA was isolated from DCs, HEV endothelial cells and non-HEV endothelial cells with an Absolutely RNA Nanoprep kit (Stratagene), and qPCR was performed as described previously<sup>30</sup>. The housekeeping gene *Ywhaz* and the endothelial cell gene *Cd31* were used as control genes for normalization. The primer sequences were as follows: *Ywhaz*, 5'-ACITTTTGGTACATTGGCTTCAA-3' and 5'-CCGCCAGGACAAACCAGTAT-3'; *Cd31*, 5'-TCCTGGGAGGTCTCCAT-3' and 5'-GAACAAGGCAGCGGGTTTA-3'; VE-cadherin, 5'-TCCTCTGCATCTCACTATCACA-3' and 5'-GTAAGTGACCAACTGCTCGTGAAT-3'; *Glycam1*, 5'-AGAATCAAGAGGCCAGGAT-3' and 5'-TGGGTCTTGTGCTCTCTCCA-3'; *FucT-VII*, 5'-CAGATGACCCCTCTAGTAGTCTGG-3' and 5'-TGCACTGTCTCTCCACAACC-3'; *GlcNAc6ST2*, 5'-GGCAAGCAGAAGGGTTAGG-3' and 5'-CTGGGAACCCAGGAACATC-3'; *Ltb*, 5'-ACCTCATAGGCGCTTGGATG-3' and 5'-ACGCTTCTCTTGGCTCGC-3'; *Lta*, 5'-CCAGGACAGCCCATCCACT-3' and 5'-GTACCAACAAGGTGAGCAGC-3'; *Light*, 5'-CGATCTCACCAGGCCAAC-3' and 5'-TCCACCAATACCTATCAAGCTG-3'; *Ccl21*, 5'-AAGGCAGTGATGGAGGGG-3' and 5'-CGGGTAAGAACAGGATTG-3'; *Cxcl13*, 5'-CATAGATCGGATTCAAGTTACG-3' and 5'-TCTTGGTCCAGATCACAACCTTCA-3'; *Icam1*, 5'-GGGAATGTGACAGGAATGT-3' and 5'-GCACCAGAATGATTATAGTCCA-3'; *Vcam1*, 5'-GGATCGCTCAAAATCGGGTGA-3' and 5'-GGTACTCGCAGCCCGTA-3'; and *Gapdh*, 5'-CCACCCAGCAAGGACACT-3' and 5'-GAAATTTGTGAGGGAGATGCTCA-3'.

**Statistical analysis.** Statistical analysis was performed using the Mann–Whitney test (for lymph node cellularity and velocity histograms) or an unpaired Student's *t*-test (for homing assays, and rolling and sticking fractions). For BMDC adoptive transfer, PBS and DTX groups were compared with a Mann–Whitney test, and DTX and DTX+BMDC groups were compared with a paired Student's *t*-test. Differences were considered statistically significant when *P* < 0.05.

31. Körner, H. et al. Distinct roles for lymphotoxin-α and tumor necrosis factor in organogenesis and spatial organization of lymphoid tissue. *Eur. J. Immunol.* **27**, 2600–2609 (1997).

# Senescence surveillance of pre-malignant hepatocytes limits liver cancer development

Tae-Won Kang<sup>1\*</sup>, Tetyana Yevsa<sup>1,2\*</sup>, Norman Woller<sup>2</sup>, Lisa Hoenicke<sup>1</sup>, Torsten Wuestefeld<sup>1,2</sup>, Daniel Dauch<sup>1</sup>, Anja Hohmeyer<sup>1,2</sup>, Marcus Gereke<sup>1</sup>, Ramona Rudalska<sup>1</sup>, Anna Potapova<sup>1</sup>, Marcus Iken<sup>3</sup>, Mihael Vucur<sup>4</sup>, Siegfried Weiss<sup>1</sup>, Mathias Heikenwalder<sup>5,6</sup>, Sadaf Khan<sup>7</sup>, Jesus Gil<sup>7</sup>, Dunja Bruder<sup>1</sup>, Michael Manns<sup>2</sup>, Peter Schirmacher<sup>8</sup>, Frank Tacke<sup>4</sup>, Michael Ott<sup>3</sup>, Tom Luedde<sup>4</sup>, Thomas Longerich<sup>8</sup>, Stefan Kubicka<sup>9</sup> & Lars Zender<sup>1,2</sup>

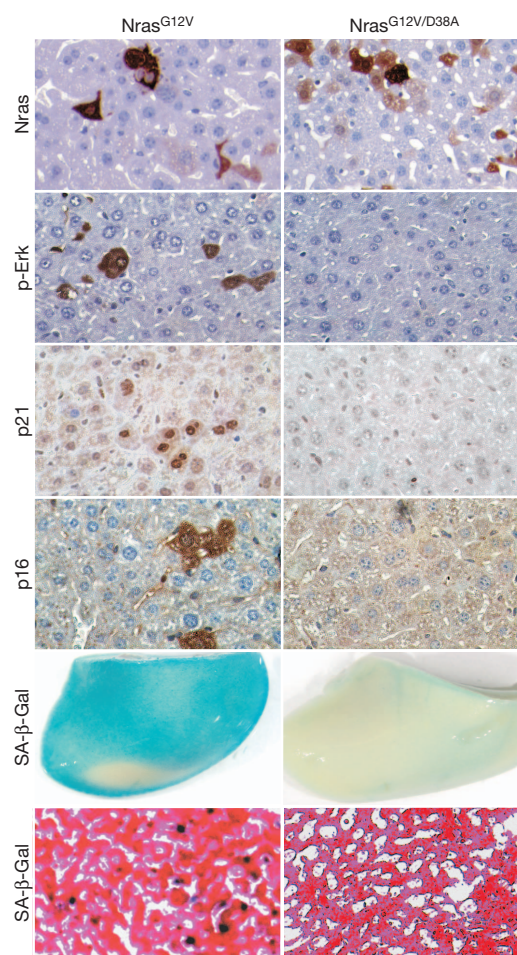
Upon the aberrant activation of oncogenes, normal cells can enter the cellular senescence program, a state of stable cell-cycle arrest, which represents an important barrier against tumour development *in vivo*<sup>1</sup>. Senescent cells communicate with their environment by secreting various cytokines and growth factors, and it was reported that this 'secretory phenotype' can have pro- as well as anti-tumorigenic effects<sup>2–5</sup>. Here we show that oncogene-induced senescence occurs in otherwise normal murine hepatocytes *in vivo*. Pre-malignant senescent hepatocytes secrete chemo- and cytokines and are subject to immune-mediated clearance (designated as 'senescence surveillance'), which depends on an intact CD4<sup>+</sup> T-cell-mediated adaptive immune response. Impaired immune surveillance of pre-malignant senescent hepatocytes results in the development of murine hepatocellular carcinomas (HCCs), thus showing that senescence surveillance is important for tumour suppression *in vivo*. In accordance with these observations, ras-specific Th1 lymphocytes could be detected in mice, in which oncogene-induced senescence had been triggered by hepatic expression of *Nras*<sup>G12V</sup>. We also found that CD4<sup>+</sup> T cells require monocytes/macrophages to execute the clearance of senescent hepatocytes. Our study indicates that senescence surveillance represents an important extrinsic component of the senescence anti-tumour barrier, and illustrates how the cellular senescence program is involved in tumour immune surveillance by mounting specific immune responses against antigens expressed in pre-malignant senescent cells.

To investigate whether oncogene-induced senescence occurs in the mouse liver, we used hydrodynamic injection<sup>6</sup> to stably deliver transposable elements<sup>7</sup> expressing oncogenic *Nras* (*Nras*<sup>G12V</sup>) into hepatocytes. Transposable elements encoding an effector loop mutant (*Nras*<sup>G12V/D38A</sup>), incapable of signalling to downstream pathways<sup>8</sup>, served as a control (Supplementary Fig. 1). Although *Nras* immunostainings revealed efficient transposon delivery in both groups (Fig. 1 and Supplementary Fig. 2), activated MAPK signalling (p-ERK immunostaining) could only be detected in animals transduced with *Nras*<sup>G12V</sup> (Fig. 1). ERK phosphorylation was accompanied by a cellular senescence response as shown by stainings for p21, p16 and SA-β-gal (Fig. 1), and quantification of *Nras*-positive and p21- and p16-positive senescent cells revealed equal numbers, indicating that every *Nras*<sup>G12V</sup>-expressing hepatocyte had entered the senescence program (Supplementary Fig. 3).

Microscopic examination of mouse livers harbouring pre-malignant senescent hepatocytes showed an inflammatory reaction with large clusters of immune cells surrounding morphologically altered, senescent hepatocytes (Fig. 2a). Similar findings were made when

senescence was triggered by Cre-recombinase-mediated intrahepatic activation of an endogenous oncogenic *Kras* allele (*Kras*<sup>G12D</sup>)<sup>9</sup> (Supplementary Fig. 4).

To address whether the observed immune cell infiltration in oncogene-induced senescence mouse livers could be due to secreted cytokines, lysates from *Nras*<sup>G12V</sup>- or *Nras*<sup>G12V/D38A</sup>-transduced livers

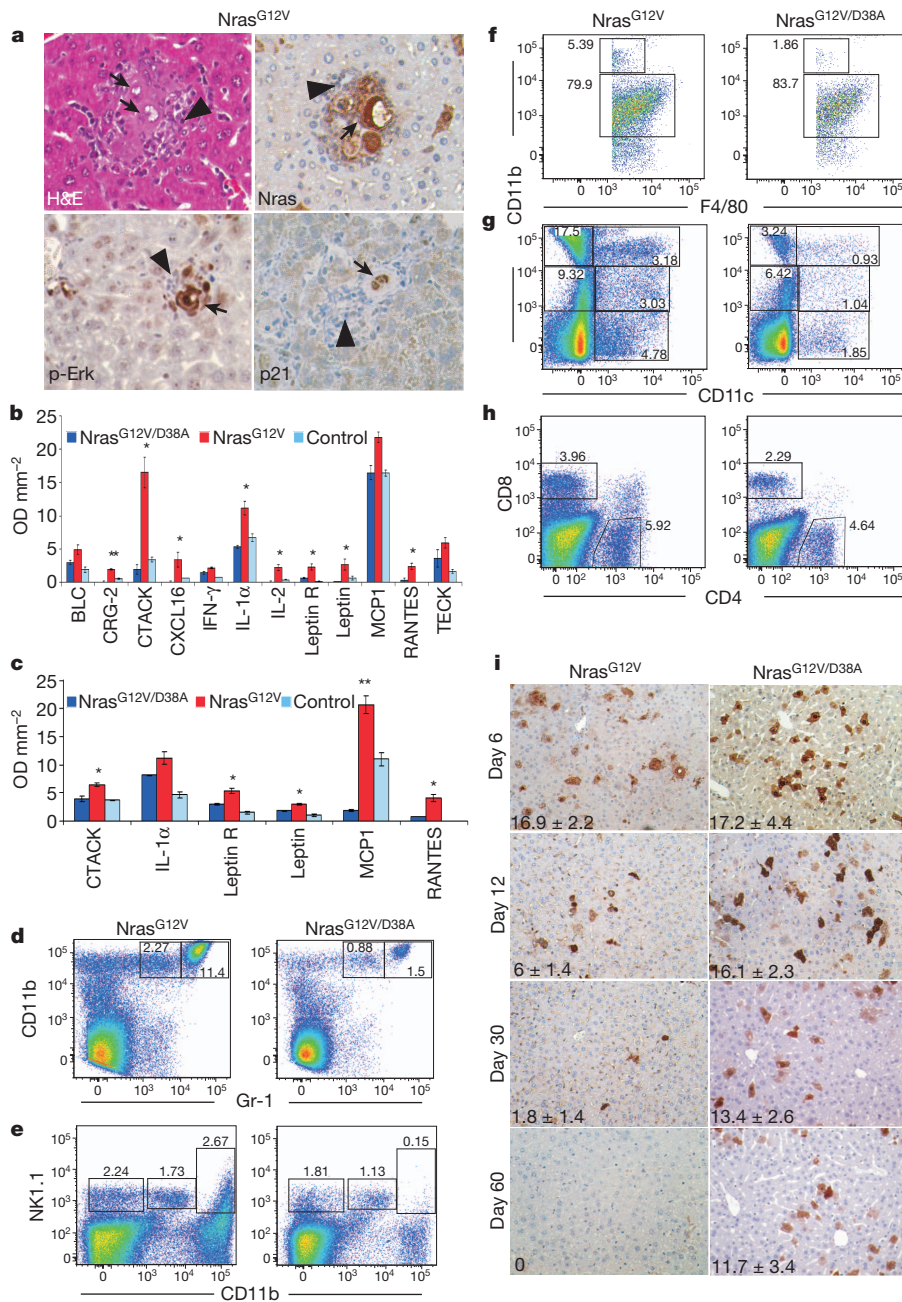


**Figure 1 | Intrahepatic expression of oncogenic *Nras*<sup>G12V</sup> induces cellular senescence.** Stainings for *Nras*, p-Erk, p21, p16 and SA-β-gal (×400) were performed on liver sections or resected liver lobes (SA-β-gal wholemount staining) 12 days after delivery of transposon constructs (see also Supplementary Fig. 1).

<sup>1</sup>Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany. <sup>2</sup>Hannover Medical School, Dept. of Gastroenterology, Hepatology and Endocrinology, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany. <sup>3</sup>Twincore Centre for Experimental and Clinical Infection Research, Feodor-Lynen Strasse 7, 30625 Hannover, Germany. <sup>4</sup>Department of Internal Medicine III, University Hospital Aachen, 52074 Aachen, Germany. <sup>5</sup>Department of Pathology, University Hospital Zurich, CH 8091 Zurich, Switzerland. <sup>6</sup>Institute of Virology, Helmholtz Centre Munich, Technical University Munich, 81675 Munich, Germany. <sup>7</sup>MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, Hammersmith Campus, London W12 0NN, UK. <sup>8</sup>University of Heidelberg, Institute of Pathology, Im Neuenheimer Feld 220/21, 69120 Heidelberg, Germany. <sup>9</sup>Clinic for Internal Medicine, District Hospital Reutlingen, Steinenbergstrasse 31, 72764 Reutlingen, Germany.

\*These authors contributed equally to this work.





**Figure 2 | Pre-malignant senescent hepatocytes are cleared by liver-infiltrating immune cells.**

**a**, Immune cell clusters were found in close proximity to pre-malignant senescent hepatocytes as shown by hematoxylin and eosin, *Nras*, p-Erk and p21 stainings ( $\times 400$ ). Arrows, senescent hepatocytes; arrowheads, infiltrating immune cells. **b, c**, Oncogene-induced senescence in mouse hepatocytes leads to an increased expression of cyto-/chemokines as measured by cytokine arrays on whole liver lysates (**b**) or purified hepatocytes (**c**). OD, optical density. \*\**P* < 0.0048 and \**P* < 0.0107, each *Nras*<sup>G12V</sup> compared with *Nras*<sup>G12V/D38A</sup>. **d-h**, Immunophenotyping and quantification of liver-infiltrating immune cells by flow cytometry: neutrophils/monocytes (**d**), NK cells (**e**), macrophages (**f**), dendritic cells (**g**) and CD4<sup>+</sup> and CD8<sup>+</sup> lymphocytes (**h**) (representative data from three independent experiments). **i**, Quantification of *Nras*-positive cells ( $\times 200$ ) on liver sections collected from mice transduced with *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> at the indicated time points. Values represent mean  $\pm$  s.d. (*n* = 4).

or purified hepatocytes thereof were analysed by cytokine arrays, which revealed that several cytokines were directly secreted by senescent hepatocytes (CTACK, interleukin-1 $\alpha$  (IL-1 $\alpha$ ), leptin R, leptin, MCP1 and RANTES), whereas others seemed to be secreted by liver-infiltrating immune cells (CRG-2, interferon- $\gamma$  (IFN- $\gamma$ ), IL-2, TECK) (Fig. 2b, c).

Using flow cytometry analyses on liver single-cell suspensions, we found that the inflammatory reaction in senescent mouse livers was composed of neutrophils (CD11b<sup>+</sup>/Gr-1<sup>high</sup>), monocytes (CD11b<sup>+</sup>/Gr-1<sup>int</sup>) (Fig. 2d), natural killer (NK) cells (NK1.1<sup>+</sup>/CD11b<sup>+</sup>) (Fig. 2e), macrophages (CD11b<sup>high</sup>/F4/80<sup>+</sup>) (Fig. 2f) and CD11c<sup>+</sup> dendritic cells (Fig. 2g). Furthermore, an increased influx of CD8<sup>+</sup> as well as CD4<sup>+</sup> T cells was detected (Fig. 2h). An increased number of CD4<sup>+</sup> T lymphocytes and dendritic cells could also be detected in portal lymph nodes of mice injected with *Nras*<sup>G12V</sup> (Supplementary Fig. 5a, b). Immunohistochemical analyses confirmed that neutrophils, lymphocytes and monocytes/macrophages were found in immune cell clusters encircling hepatocytes expressing *Nras*<sup>G12V</sup> (Supplementary Fig. 6).

We next sought to investigate whether liver-infiltrating immune cells mediate the clearance of senescent hepatocytes. Although 6 days after *in vivo* delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> similar numbers of *Nras*-expressing cells were found in both groups, time-course analysis revealed a rapid loss of *Nras*-positive cells in *Nras*<sup>G12V</sup>-transduced livers, with only  $6 \pm 1.4\%$  of *Nras*-positive cells remaining at day 12 and no detectable *Nras*-positive cells after 60 days, whereas no such change was observed in *Nras*<sup>G12V/D38A</sup>-transduced livers (Fig. 2i). Examination of livers by TdT-mediated dUTP nick end labelling (TUNEL) staining did not reveal significant differences, thus excluding the possibility that an increased cell death in *Nras*<sup>G12V</sup>-transduced livers accounts for the decrease in *Nras*-positive cells (Supplementary Fig. 7).

To delineate which immune cells are mediating the clearance of senescent hepatocytes, we took advantage of syngeneic mice harbouring genetically defined immune defects: C.B-17 wild-type (WT) mice with an intact immunity, C.B-17 SCID mice harbouring an impaired adaptive immune response and C.B-17 SCID/beige mice, which furthermore exhibit defects in NK-cell and macrophage function

and generally in phagocytosis (Fig. 3a). Transposable elements encoding *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> were intrahepatically delivered into these mice, and numbers of *Nras*-positive and senescent (p16- and p21-positive) hepatocytes were determined. As expected, C.B-17 WT mice showed an accented reduction of *Nras*-positive and senescent cells over time, with a complete loss of *Nras*-positive cells after 60 days (Fig. 3b–e). In strong contrast, no significant reduction in the number of *Nras*-positive cells was found in C.B-17 SCID mice after intrahepatic *Nras*<sup>G12V</sup> delivery (Fig. 3b–e), thus suggesting that a defective adaptive immunity abrogates immune surveillance of pre-malignant senescent hepatocytes. Interestingly, significantly less p21- and p16 positive than *Nras*-positive hepatocytes could be detected at day 30 and 60 in C.B-17 SCID mice, suggesting that some *Nras* expressing hepatocytes at this time point have escaped the senescence program (Fig. 3b–e and Supplementary Fig. 8).

To explore the long-term consequences of defective senescence surveillance, we followed up C.B-17, C.B-17 SCID and C.B-17 SCID/beige mice after stable *in vivo* delivery of oncogenic *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup>. We found dysplastic nodules, a bona fide precursor lesion for the development of HCC, as early as after 3 months

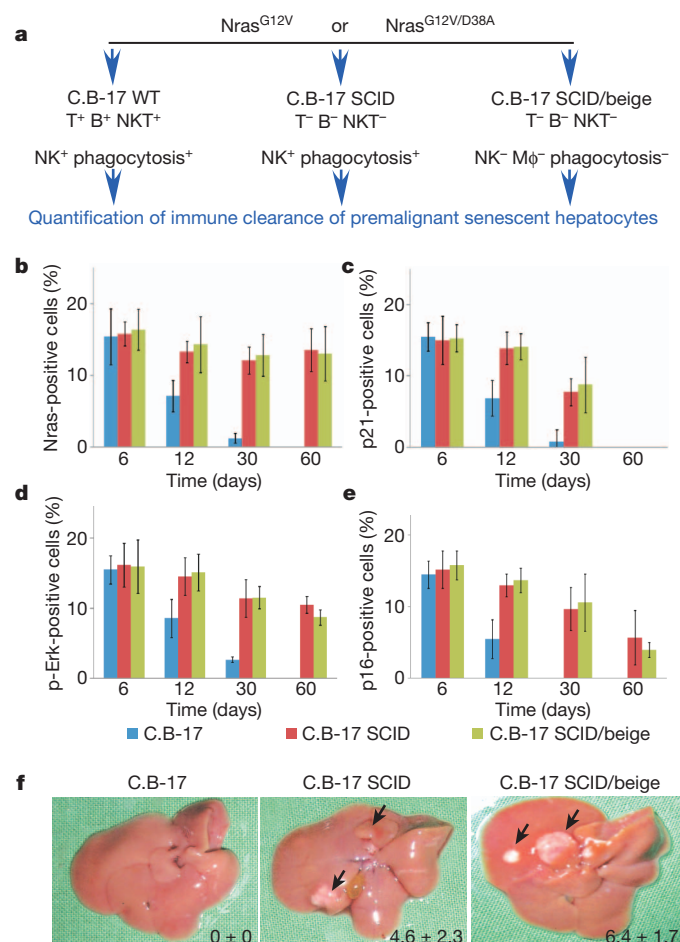
(Supplementary Fig. 9a), followed by HCCs as early as 4 months, after intrahepatic delivery of oncogenic *Nras* in C.B-17 SCID and C.B-17 SCID/beige mice (Supplementary Fig. 9b and data not shown). Although after 7 months there was still no tumour growth in the control group (*Nras*<sup>G12V/D38A</sup>-transduced mice) (Supplementary Fig. 10) and in *Nras*<sup>G12V</sup>-expressing immunocompetent C.B-17 mice (Fig. 3f), severe intrahepatic tumour growth could be observed in *Nras*<sup>G12V</sup>-transduced C.B-17 SCID as well as C.B-17 SCID/beige mice (Fig. 3f and Supplementary Fig. 11), suggesting that an impaired immune surveillance of pre-malignant senescent hepatocytes results in HCC development.

The SCID immune defect comprises defects in T lymphocyte, B lymphocyte and NKT cell function<sup>10</sup>. To address the role of these cell types in senescence surveillance, we next analysed senescence surveillance in *Cd1d*<sup>-/-</sup> (lacking NKT cells), *CD8*<sup>-/-</sup> and *CD4*<sup>-/-</sup> mice as well as in immunocompetent mice where *CD4*<sup>+</sup> and *CD8*<sup>+</sup> T cells had been depleted by antibodies. While *Cd1d*<sup>-/-</sup> and *CD8*<sup>-/-</sup> mice and those with antibody-mediated depletion of *CD8*<sup>+</sup> T cells displayed normal or nearly normal senescence surveillance (Supplementary Fig. 12 and Fig. 4a), a completely blunted senescence surveillance was observed in *CD4*<sup>-/-</sup> mice and mice of different H-2 haplotypes, in which *CD4*<sup>+</sup> T lymphocytes had been depleted by antibodies (Fig. 4a and Supplementary Fig. 13). *CD4*<sup>+</sup> T-cell-dependent surveillance of pre-malignant senescent hepatocytes was also found in a mouse model, wherein liver-specific activation of oncogenic ras was accomplished by a tamoxifen-inducible cre-recombinase<sup>11,12</sup> (Supplementary Fig. 14).

In line with these results, we observed that stable delivery of oncogenic *Nras*<sup>G12V</sup> resulted in massive intrahepatic tumour growth in *CD4*<sup>-/-</sup> mice (Fig. 4b and Supplementary Fig. 15), whereas all arising tumours were found to harbour the *Nras*<sup>G12V</sup> encoding transposon (Supplementary Fig. 16a, b). Western blotting of tumour protein lysates against p19<sup>Arf</sup> suggested reduced p19<sup>Arf</sup> expression as a potential mechanism that had allowed *Nras*<sup>G12V</sup>-expressing hepatocytes to escape from the senescence arrest (Supplementary Fig. 17).

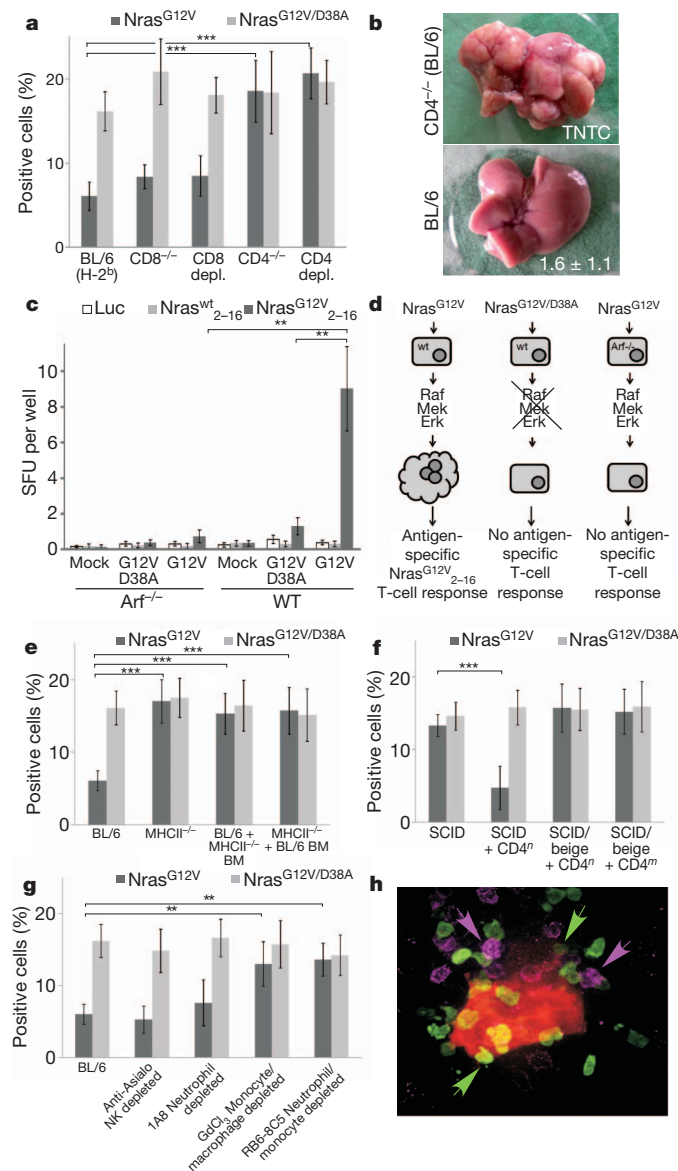
IFN- $\gamma$  enzyme-linked immunospot (ELISPOT) assays with mutant *Nras*-specific 15-mer peptides revealed antigen-specific T cells in mice harbouring senescent hepatocytes (*Nras*<sup>G12V</sup>) but not in mice expressing *Nras*<sup>G12V/D38A</sup> (Fig. 4c and Supplementary Fig. 18). To probe whether the induction of a mutant ras-specific immune response is indeed dependent on execution of the senescence program, we repeated our experiments in p19<sup>Arf</sup><sup>-/-</sup> mice, in which the senescence program is genetically disabled<sup>13</sup>. Intrahepatic delivery of *Nras*<sup>G12V</sup> into p19<sup>Arf</sup><sup>-/-</sup> mice did not trigger ras-specific IFN- $\gamma$  production (Fig. 4c, d); rather, it resulted in the rapid onset of invasive HCC, which was independent of the *CD4*<sup>+</sup> T-cell status (Supplementary Fig. 19a, b). Antigen-specific *CD4*<sup>+</sup> T-cell responses were confirmed in mice that had been transduced with transposons co-expressing *Nras*<sup>G12V</sup> and haemagglutinin (HA), as ELISPOT assays with splenocytes of these mice indicated a strong, *CD4*<sup>+</sup> T-cell-specific immune response against HA (Supplementary Fig. 20a). No significant IL-4 and IL-17 responses could be measured in these experiments, thus suggesting that the observed response represents a Th1-polarized *CD4*<sup>+</sup> T-cell response (Supplementary Fig. 20a–c).

We next sought to explore whether senescence surveillance also occurs in inflammation-based HCC mouse models. For this, we analysed *CD4*<sup>+</sup> T-cell-dependent senescence surveillance and tumour development in mice with liver-specific ablation of MAP3-kinase transforming growth factor (TGF)- $\beta$ -activated kinase 1 (Tak1)<sup>14</sup>. Antibody-mediated depletion of *CD4*<sup>+</sup> T cells resulted in a strong accumulation of senescent hepatocytes in *Tak1*<sup>-/-</sup> mice, which correlated with increased liver tumour development already at 3 months of age (Supplementary Fig. 21a–c). Furthermore, we quantified senescent hepatocytes in livers of *Rag1*<sup>-/-</sup> or syngeneic WT mice, which had been subjected to the diethylnitrosamine (DEN) liver carcinogen<sup>15</sup>. Interestingly, *Rag1*<sup>-/-</sup> mice showed an accumulation



**Figure 3 | Impaired senescence surveillance results in liver cancer development.** **a**, Experimental setup and description of immune status of C.B-17 WT, C.B-17 SCID and C.B-17 SCID/beige mice. **b–e**, Quantification of *Nras*-positive (b), p21- (c), p-Erk- (d) and p16-positive cells (e) on liver sections from indicated mice after stable intrahepatic delivery of *Nras*<sup>G12V</sup>. Values represent mean ± s.d. (*n* = 4). **f**, Impaired immune surveillance of pre-malignant senescent hepatocytes results in the development of HCC in C.B-17 SCID and C.B-17 SCID/beige mice that had been transduced with *Nras*<sup>G12V</sup>. Shown are representative photographs of explanted livers (*n* = 10). Arrows, intrahepatic tumour nodules. Values represent mean ± s.d. of macroscopically detectable tumour nodules bigger than 3 mm in diameter.





of senescent hepatocytes 24 weeks after DEN administration (Supplementary Fig. 22a, b), which was associated with an accelerated and increased tumorigenesis at later time points (data not shown, and F. Tacke and C. Schneider, unpublished observations). As the *Rag1*<sup>-/-</sup> gene defect comprises both CD4 and CD8 deficiency, further work is needed to delineate the role of CD4<sup>+</sup> and CD8<sup>+</sup> T cells in this model.

It has been described that hepatocytes can directly present antigens through major histocompatibility complex class II (MHCII); however, the significance of this process is still under debate. We studied antigen-specific CD4<sup>+</sup> T-cell activation in mice with liver- or bone-marrow-specific MHCII deficiency and found that CD4<sup>+</sup> T-cell activation depends on the presence of professional antigen presenting cells (APCs) (Supplementary Fig. 23 and supplementary data description). These findings were corroborated by *in vitro* co-incubation assays of senescent, *Nras*<sup>G12V</sup>-HA-expressing hepatocytes with TCR-HA CD4<sup>+</sup> T cells, confirming that proliferation was dependent on the presence of MHCII-matched APCs and cytokine secretion from senescent hepatocytes and immune cells (Supplementary Fig. 24a, b). Furthermore, functional *in vivo* experiments revealed a crucial role for M1 cytokines, in particular IL-1 $\alpha$ , in senescence surveillance (Supplementary Fig. 25), whereas cytokine secretion seems to be p38 MAPK dependent (Supplementary Fig. 26a, b).

**Figure 4 | Immune surveillance of pre-malignant senescent hepatocytes is orchestrated by antigen-specific CD4<sup>+</sup> T cells.** **a**, Quantification of *Nras*-positive cells on liver sections from WT (C57BL/6), CD4<sup>-/-</sup>, CD8<sup>-/-</sup> mice or CD4- or CD8-depleted WT mice 12 days after intrahepatic delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> ( $n = 5$ ). Values represent mean  $\pm$  s.d. \*\*\* $P < 0.0004$  (*Nras*<sup>G12V</sup>, C57BL/6 compared with CD4<sup>-/-</sup>), \*\*\* $P < 0.0001$  (*Nras*<sup>G12V</sup>, C57BL/6 compared with CD4 depleted). **b**, Impaired senescence surveillance results in the development of HCCs in CD4<sup>-/-</sup> mice 7 months after intrahepatic delivery of *Nras*<sup>G12V</sup> ( $n = 10$ ). Values represent mean  $\pm$  s.d. of macroscopically detectable tumour nodules bigger than 3 mm in diameter (TNTC, too numerous to count). **c**, IFN- $\gamma$  ELISPOT assay performed on splenocytes isolated from WT or p19<sup>Arf</sup><sup>-/-</sup> mice 20 days after intrahepatic delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup>. Results are represented as spot-forming units (SFU) per well  $\pm$  s.e.m. ( $n = 8$ ; \*\* $P < 0.0026$  (*Nras*<sup>G12V</sup>, C57BL/6/*Nras*<sup>G12V</sup> compared with p19<sup>Arf</sup><sup>-/-</sup>/*Nras*<sup>G12V</sup>), \*\* $P < 0.0065$  (*Nras*<sup>G12V</sup>, C57BL/6/*Nras*<sup>G12V</sup> compared with C57BL/6 *Nras*<sup>G12V/D38A</sup>)). **d**, Representation of the pathway: induction of an antigen-specific immune response is dependent on the execution of the senescence program. **e**, Quantification of *Nras*-positive cells 12 days after intrahepatic delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> on liver sections from C57BL/6 WT, MHCII<sup>-/-</sup> mice and chimaeras as described in Supplementary Fig. 23a–d ( $n = 5$ ; \*\*\* $P < 0.0003$ ). **f**, Quantification of *Nras*-positive cells on liver sections from C.B-17 SCID mice, C.B-17 SCID mice adoptively transferred with syngeneic CD4<sup>+</sup> T cells, and C.B-17 SCID/beige mice adoptively transferred with syngeneic CD4<sup>+</sup> T cells (CD4<sup>n</sup>, naive; CD4<sup>m</sup>, memory) 12 days after intrahepatic delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> ( $n = 5$ ; \*\*\* $P < 0.0002$ ). **g**, Quantification of *Nras*-positive cells on liver sections from untreated, NK cell-depleted, Ly-6G-depleted (clone 1A8), GdCl<sub>3</sub>-depleted and Ly-6G/Ly-6C (clone RB6-8C5)-depleted C57BL/6 mice 12 days after intrahepatic delivery of *Nras*<sup>G12V</sup> or *Nras*<sup>G12V/D38A</sup> ( $n = 5$ ). Values represent mean  $\pm$  s.d. (\*\* $P < 0.0054$ ). **h**, Confocal microscopy on lys-EGFP transgenic mouse livers (EGFP<sup>+</sup> monocytes/macrophages and neutrophils). Senescent hepatocytes stain orange (co-expression of oncogenic *Nras*<sup>G12V</sup> with TagRFP). Green arrows, EGFP-positive monocytes/macrophages and neutrophils; pink arrows, APC-labelled lymphocytes ( $\times 600$ ).

Interestingly, we could detect upregulation of MHCII on a subfraction of senescent hepatocytes and concomitant upregulation of co-stimulatory CD86 on some cells (Supplementary Fig. 24c). It is noteworthy that, although not sufficient to induce proliferation of naive CD4<sup>+</sup> T cells, hepatic MHCII expression appears to be necessary to allow a directed killing of pre-malignant senescent hepatocytes as senescence surveillance was significantly impaired in mice with liver-specific MHCII deficiency (Fig. 4e).

Finally, we addressed whether senescent hepatocytes are directly killed by cytotoxic CD4<sup>+</sup> T cells<sup>16</sup> or whether CD4<sup>+</sup> T cells require innate immune cells to execute the killing. We found that adoptive transfer of syngeneic CD4<sup>+</sup> T cells into *Nras*<sup>G12V</sup>-injected C.B-17 SCID mice restored senescence surveillance (Fig. 4f) and suppressed tumour development in these mice (Supplementary Fig. 27). However, even the adoptive transfer of 'memory' CD4<sup>+</sup> T cells (obtained from immunocompetent C.B-17 mice with hepatic *Nras*<sup>G12V</sup> expression) did not restore senescence surveillance in C.B-17 SCID/beige mice (Fig. 4f), suggesting that CD4<sup>+</sup> T cells act in a T-helper cell manner and require innate immune cells to kill pre-malignant senescent cells. In accordance with these results, we found that gadolinium treatment of mice, which depletes macrophages and affects monocytes<sup>17</sup>, dramatically reduced senescence surveillance (Fig. 4g). In contrast, selective depletion of neutrophils (Ly-6G antibody) only marginally affected senescence surveillance, and antibody-mediated depletion of NK cells showed no effect at all (Fig. 4g). To distinguish whether Kupffer cells as resident liver macrophages execute the killing for CD4<sup>+</sup> T cells or whether freshly infiltrated monocytes and replenished macrophages thereof are needed, we took advantage of the RB6-8C5 (Ly-6G/Ly-6C) antibody, which depletes neutrophils and monocytes, but does not affect tissue-resident macrophages<sup>18</sup>. Interestingly, monocyte depletion by RB6-8C5 completely abrogated senescence surveillance (Fig. 4g), thus pinpointing monocytes and probably freshly replenished macrophages thereof as effector cells. The interaction between CD4<sup>+</sup> T cells and monocytes/macrophages was



confirmed by confocal microscopy on Lys-enhanced green fluorescent protein (EGFP) transgenic mouse livers, where Lys-EGFP-positive cells (EGFP<sup>+</sup> macrophages, monocytes and neutrophils) and APC-labelled lymphocytes were found in close proximity to senescent Nras<sup>G12V</sup>/TaqrFP-co-expressing hepatocytes (Fig. 4h).

Finally, to explore whether senescence surveillance occurs in human livers, we quantified senescent cells in patients infected with hepatitis C virus (HCV) with or without concomitant human immunodeficiency virus (HIV) infection (impaired CD4<sup>+</sup> T-cell function) and found that senescent (p16-positive) cells significantly accumulated in livers of patients with HIV (Supplementary Fig. 28a, b). Furthermore, we analysed a unique cohort of patients who underwent two subsequent liver transplantations due to HCV-induced liver cirrhosis with a comparable grade of liver damage. We quantified senescent cells in the patients' own explant livers (no immunosuppressive therapy, 'HCV-Tx') and in the explanted transplant liver after the development of re-cirrhosis (patient has been under immunosuppressive therapy, 'HCV-Re-Tx') and found that senescent hepatocytes significantly accumulate in livers under immunosuppressive therapy (Supplementary Fig. 28a, b), thus providing the first evidence for immune surveillance of senescent hepatocytes in human livers and one possible explanation for the increased rates of hepatocellular carcinoma found in HIV-infected patients or those under immunosuppressive therapy (reviewed in ref. 19).

Taken together, our results establish senescence surveillance as an important extrinsic component of the senescence anti-tumour barrier. Our data suggest that pre-malignant hepatocytes can escape the senescence arrest and contribute to development of liver cancer, thus emphasizing the importance of senescence surveillance. Our results on senescence escape of pre-malignant hepatocytes are in line with published reports demonstrating that senescence escape can occur *in vitro*<sup>20,21</sup>, although it is formally possible that escaping cells initiate the senescence program (and thus display senescence markers) but never achieve the final state. Furthermore, it will be important to characterize whether factors secreted from pre-malignant senescent hepatocytes also contribute to the oncogenic transformation of neighbouring cells.

Recent work has shown that innate immune cells alone are sufficient to mediate the clearance of ras-driven liver carcinomas, wherein senescence was triggered by reactivation of endogenous p53<sup>3</sup>. In contrast, here we show that senescence surveillance of pre-malignant hepatocytes is orchestrated by antigen-specific CD4<sup>+</sup> T cells.

We anticipate that strategies specifically harnessing the identified mechanism of senescence surveillance hold great promise for improving cancer prevention and therapy. Furthermore, use of the senescence program to trigger or enhance antigen-specific immune responses will very probably find broad applications to improve vaccination strategies.

## METHODS SUMMARY

For transposon-mediated intrahepatic gene transfer mice received a 5:1 molar ratio of transposon- to transposase encoding vector (30 µg total DNA). DNA was suspended in saline solution at a final volume of 10% of the animals' body weight and administered by hydrodynamic tail vein injection. Immunohistochemical analyses were performed on paraffin-embedded liver sections (citrate buffer antigen retrieval) using the indicated antibodies with standard protocols.

Received 14 July 2010; accepted 30 September 2011.

Published online 9 November 2011.

1. Narita, M. & Lowe, S. W. Senescence comes of age. *Nature Med.* **11**, 920–922 (2005).
2. Krtolica, A., Parrinello, S., Lockett, S., Desprez, P. Y. & Campisi, J. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proc. Natl Acad. Sci. USA* **98**, 12072–12077 (2001).

3. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **455**, 656–660 (2007).
4. Acosta, J. C. *et al.* Chemokine signaling via the CXCR2 receptor reinforces senescence. *Cell* **133**, 1006–1018 (2008).
5. Kuilman, T. *et al.* Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* **133**, 1019–1031 (2008).
6. Hickman, M. A. *et al.* Gene expression following direct injection of DNA into liver. *Hum. Gene Ther.* **5**, 1477–1483 (1994).
7. Carlson, C. M., Frandsen, J. L., Kirchhof, N., Mlvor, R. S. & Largaespa, D. A. Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. *Proc. Natl Acad. Sci. USA* **102**, 17059–17064 (2005).
8. Khwaja, A., Rodriguez-Vician, P., Wennstrom, S., Warne, P. H. & Downward, J. Matrix adhesion and Ras transformation both activate a phosphoinositide 3-OH kinase and protein kinase B/Akt cellular survival pathway. *EMBO J.* **16**, 2783–2793 (1997).
9. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
10. Thomsen, M., Galvani, S., Canivet, C., Kamar, N. & Bohler, T. Reconstitution of immunodeficient SCID/beige mice with human cells: applications in preclinical studies. *Toxicology* **246**, 18–23 (2008).
11. Tannour-Louet, M., Porteu, A., Vaulont, S., Kahn, A. & Vasseur-Cognet, M. A tamoxifen-inducible chimeric Cre recombinase specifically effective in the fetal and adult mouse liver. *Hepatology* **35**, 1072–1081 (2002).
12. Harada, N. *et al.* Hepatocarcinogenesis in mice with  $\beta$ -catenin and Ha-ras gene mutations. *Cancer Res.* **64**, 48–54 (2004).
13. Kamijo, T. *et al.* Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. *Cell* **91**, 649–659 (1997).
14. Bettermann, K. *et al.* TAK1 suppresses a NEMO-dependent but NF- $\kappa$ B-independent pathway to liver cancer. *Cancer Cell* **17**, 481–496 (2010).
15. Sarma, D. S., Rao, P. M. & Rajalakshmi, S. Liver tumour promotion by chemicals: models and mechanisms. *Cancer Surv.* **5**, 781–798 (1986).
16. Brown, D. M. Cytolytic CD4 cells: direct mediators in infectious disease and malignancy. *Cell. Immunol.* **262**, 89–95 (2010).
17. Frid, M. G. *et al.* Hypoxia-induced pulmonary vascular remodeling requires recruitment of circulating mesenchymal precursors of a monocyte/macrophage lineage. *Am. J. Pathol.* **168**, 659–669 (2006).
18. Silva, M. T. When two is better than one: macrophages and neutrophils work in concert in innate immunity as complementary and cooperative partners of a myeloid phagocyte system. *J. Leukoc. Biol.* **87**, 93–106 (2010).
19. Hensel, M. *et al.* HIV and cancer in Germany. *Dtsch. Arztebl. Int.* **108**, 117–122 (2011).
20. Sage, J., Miller, A. L., Perez-Mancera, P. A., Wysocki, J. M. & Jacks, T. Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. *Nature* **424**, 223–228 (2003).
21. Dirac, A. M. & Bernards, R. Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. *J. Biol. Chem.* **278**, 11731–11734 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Largaespa and M. Kay for providing transposon vectors and transposase encoding vectors. We thank L. Gröbe, A. Rinkel, N. Struever, H. Riedesel, the team of the Helmholtz Centre for Infection Research (HZI) animal facility, M. Rothe, K. Schulze, A. Kobold, F. Heinzmann, N. Jedicke, C. Schneider, M. Pesic, H. Klimek and A. Samuels for technical assistance and assistance with animal work, and F. Alves, S. Kimmina, C. Dullin and S. Greco for assistance with flat panel-volumetric computer tomography. We thank the tissue bank of the National Center for Tumor Diseases Heidelberg for providing liver explant tissues. We thank S. Lowe, H. Tillmann, F. Greten and members of the Lowe and Zender laboratory for advice and discussions. This work was supported by the Helmholtz Association of German Research Centres (VH-NG-424 to L.Z.), the German Research Foundation, DFG (Emmy Noether Programme ZE 545/2-1 to L.Z., SFB/TRR77 and the 'Rebirth' Cluster of Excellence), the Wilhelm Sander Stiftung, the Bear Necessities Pediatric Cancer Foundation, the Federal German Ministry for Education and Research (BMBF) (ARCHES AWARD to L.Z.) and the European Commission (project 'Heptromic'). L.Z. holds an adjunct assistant Professorship with the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. This work is dedicated to J. Wehland.

**Author Contributions** L.Z. designed and guided the research and wrote the manuscript. T.-W.K., T.Y. and N.W. conducted experiments and contributed to research design and manuscript preparation. L.H., T.W., D.D., A.H., M.G., R.R., A.P., M.I., M.V., S.W., M.H., S.K., J.G., F.T., To.L., D.B. M.M., M.O. and S.K. contributed to research design and/or conducted experiments. P.S. and Th.L. performed histopathological analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.Z. (Lars.Zender@helmholtz-hzi.de).